UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

The Role of Nanog Regulating Pluripotency: A Single Cell Approach

Gonçalo Filipe Ferreira Fernandes

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada por:
Domingos Henrique
Maria Gabriela Rodrigues

2017

Resumo

As células estaminais embrionárias (CEE) são células derivadas de células do botão embrionário ou epiblasto de blastocistos, podendo ser mantidas em cultura *in vitro* indefinidamente. Estas células mantêm características semelhantes às células das quais são derivadas, sendo pluripotentes e tendo capacidade de autorrenovação, ou seja, têm a capacidade de diferenciarem em todas as células existentes num organismo adulto, e a cada divisão são capazes de gerar pelo menos uma célula-filha com as mesmas características que a célula original. Devido a estas características únicas, se estas células forem colocadas num blastocisto, contribuem para o normal desenvolvimento do embrião, que se torna uma quimera, com células provenientes do blastocisto e das células estaminais embrionárias.

Devido a estas razões, as CEE têm-se tornado uma ótima ferramenta para estudar o desenvolvimento embrionário. A recente descoberta de que células diferenciadas podem ser induzidas a recuperar o fenótipo de pluripotência deu um novo fôlego ao estudo dos mecanismos necessários para a indução de pluripotência e de diferenciação. Como estas células têm a capacidade de se diferenciarem em todos os tipos celulares existentes num organismo adulto, também têm sido muito estudadas para futuras aplicações em medicina regenerativa.

Estas características especiais das células estaminais, pluripotência e autorrenovação, são mantidas por uma rede de fatores de transcrição, centrada em três elementos: Oct4, Sox2 e Nanog. Estes fatores de transcrição atuam em conjunto, regulando a sua própria expressão e a de outros genes envolvidos na manutenção da pluripotência, e reprimindo a expressão de genes envolvidos em diferenciação. Enquanto que Oct4 e Sox2 são homogeneamente expressos nas condições normais de cultura de células pluripotentes, Nanog é expresso de forma heterogénea, sendo muito expresso em algumas células e pouco expresso noutras. Estas células que apresentam baixa expressão de Nanog também apresentam expressão de genes normalmente associados com diferenciação. O facto destes genes serem expressos maioritariamente em células com baixa expressão de Nanog levou à criação de uma hipótese em que o Nanog poderá ser um regulador destes genes, reprimindo-os quando é expresso.

O ambiente de cromatina é um dos principais reguladores da expressão genética, tendo alguns estudos desvendado uma ligação entre o Nanog e enzimas modificadoras da cromatina, como Tet1 (ten-eleven translocation 1) e Ezh2. Tet1 é uma enzima responsável pela demetilação do DNA, sendo o seu substrato as citosinas metiladas (5mC) do DNA. Ezh2 é o componente enzimático do complexo PRC2 (polycomb repressive complex 2), responsável pela metilação da lisina 27 da histona H3 (H3K27). Demetilação da 5mC é normalmente associada a ativação da expressão genética, e metilação da H3K27 associada a repressão genética. No entanto, também já foi descrito que o PRC2 se liga a regiões hipometiladas do genoma, como aquelas que são criadas pela Tet1.

Neste projeto, é proposto um mecanismo de ação do Nanog na regulação dos genes envolvidos em diferenciação, em que o Nanog recruta Tet1 para as regiões promotoras desses genes, permitindo a sua demetilação (5mC para 5hmC). Estes promotores hipometilados são reconhecidos pelo PRC2, que metila as histonas H3 que estão na vizinhança, no resíduo K27.

Este mecanismo foi testado através da utilização de moduladores químicos, que afetam a atividade da Tet1 e do PRC2. A atividade da Tet1 é estimulada por ácido ascórbico, também conhecido por vitamina C, e a atividade do PRC2 é reprimida por GSK343, um composto inibidor da atividade da Ezh2, a subunidade catalítica do PRC2 responsável pela metilação da histona H3. Assim, a utilização do ácido ascórbico deverá levar à repressão dos genes envolvidos na diferenciação através da estimulação da atividade da Tet1, levando ao aumento de regiões hipometiladas, à ligação do PRC2 e consequente

metilação de H3K27. Por outro lado, a repressão do PRC2 deverá causar o aumento de expressão dos genes envolvidos em diferenciação devido à não metilação de H3K27.

A heterogeneidade da expressão do *Nanog* e, por comparação, a homogeneidade do *Sox2* foram analisadas através de uma técnica que permite a quantificação do número exato de moléculas de mRNA existentes numa célula, sendo que a análise de muitos células permite medir as características da população. Os nossos resultados mostram que o *Nanog* apresenta uma grande heterogeneidade ao nível da expressão do mRNA, com células com um reduzido número de transcritos e células com elevado número de mRNAs. Por outro lado, observa-se uma grande frequência de células com um número médio de transcritos de *Sox2*, e poucas células com maior ou menor número de transcritos. Entre estes dois genes existe uma correlação positiva, indicando que a maioria das células com alta expressão de um dos genes também tem alta expressão do outro, à exceção de 30% de células que apresentam alta expressão de *Sox2* e baixa expressão de *Nanog*. Estas células correspondem às que poderão estar a explorar a pluripotência através da expressão de genes envolvidos em diferenciação. A expressão de alguns destes genes foi também analisada, com os resultados a mostrarem que a maioria das células os expressa a baixos níveis, existindo alta expressão em apenas um pequeno número de células, maioritariamente células com baixa expressão de *Nanog*.

A expressão de *Fgf5* e *Sox3*, dois dos genes envolvidos em diferenciação e expressos em células com baixa expressão de *Nanog*, foi também analisada em células de subpopulações puras com alta e baixa expressão de Nanog, sorteadas com base na presença duma proteína fluorescente, cujos níveis de expressão mimetizam a presença de Nanog. Nestas populações é possível observar uma clara distinção entre os dois estados de expressão do Nanog, em que uma população expressa altos níveis de *Nanog* e baixos níveis de *Fgf5* e *Sox3*, enquanto que a outra população expressa baixos níveis de *Nanog* e altos níveis de *Fgf5* e *Sox3*. Os baixos níveis de expressão destes dois genes nas células com alta expressão de Nanog permitiram a definição de alta expressão de genes envolvidos em diferenciação, definindo um limiar entre baixa e alta expressão.

Tendo confirmado a heterogeneidade da expressão do Nanog e a maior expressão de genes envolvidos em diferenciação em células que apresentam baixa expressão de Nanog, o modelo proposto foi testado. O modelo foi testado numa população normal (heterogénea) e em subpopulações puras com alta e baixa expressão de Nanog (sorteadas com base na expressão de uma proteína fluorescente, que mimetiza a expressão do Nanog). Na população normal foram testados os genes *Car2*, *Crabp2*, *Fgf5* e *Sox3*, enquanto que nas experiências com as subpopulações apenas foram testados os genes *Car2* e *Sox3*.

Os resultados obtidos com o trabalho descrito nesta dissertação corroboram, através de evidências experimentais, o modelo aqui proposto para regulação de expressão de genes envolvidos em diferenciação, dando como exemplo o *Sox3*, por parte do Nanog, em cooperação com Tet1 e PRC2.

Palavras-chave: Células Estaminais Embrionárias, Nanog, Pluripotência, *Lineage-priming*, smRNA-FISH

Abstract

Embryonic stem cells (ESC) are derived from the epiblast region of blastocysts and characterized by self-renewal and pluripotency. These characteristics are maintained by the activity of the pluripotency network, at the core of which functions a trio of transcription factors, namely Oct4, Sox2 and Nanog. In cultured mouse ESCs (mESCs), Oct4 and Sox2 are homogeneously expressed, while Nanog shows a remarkable heterogeneity. Hence, it is possible to distinguish between states of low and high Nanog expression (Low- and High-Nanog, respectively), which previous work has shown to be functionally and molecularly different.

In the High-Nanog state, the pluripotency network is fully active, maintaining pluripotency and repressing differentiation, whereas mESCs at the Low-Nanog state are characterized by low level expression of genes usually involved in lineage-choice and differentiation ("priming genes"). This observation led to the proposal that Nanog might be a regulator of these priming genes, and that the observed Nanog fluctuations provide windows of opportunity within the pluripotent state, during which mESCs can be primed for lineage differentiation.

The chromatin environment is also a key regulator of pluripotency and differentiation, and Nanog has been shown to interact with chromatin modulating enzymes, like ten-eleven translocation 1 (Tet1) and members of the polycomb repressive complex (PRC2). Tet1 is an enzyme responsible for the demethylation of methylated cytosines (5mC) in DNA, and PRC2 is a complex responsible for the trimethylation of histone H3 at the lysine residue 27 (H3K27), on hypomethylated regions of DNA.

Previous work in the Henrique's laboratory led to a model to understand priming gene regulation by Nanog in mESCs. In this model, Nanog binds to regulatory regions of the priming genes and recruits Tet1, which will catalyse the conversion of 5mC into 5hmC to generate hypomethylated regions (normally in CpG islands). These regions are then recognized by PRC2, that would methylate associated H3 at K27, leading to repression of priming genes. As this regulation depends on Nanog expression, it should work when ESCs transit from the Low-Nanog primed state to High-Nanog, during the observed Nanog fluctuations.

To test the proposed model, mESCs were exposed to small molecule chemical modulators, ascorbic acid (AA) and GSK343, which interfere with Tet1 and PRC2 activities, and gene expression was measured at the single cell level by single molecule RNA-FISH. AA is known to stimulate Tet1 activity, and GSK343 is an inhibitor of Ezh2 activity, the catalytical subunit of PRC2. According to the model, AA should decrease priming gene expression by promoting CpG demethylation of priming genes, and subsequent PRC2 recruitment and H3K27 methylation. On the contrary, GSK343-mediated Ezh2 inhibition should lead to increased priming gene expression by preventing establishment of H3K27 methylation around priming genes.

The results obtained in this thesis, using *Sox3* expression as an illustrative priming gene, provide experimental evidence that supports the proposed model of priming gene regulation by Nanog, working with Tet1 and PRC2.

Keywords: Embryonic Stem Cells, Nanog, Pluripotency, Lineage-priming, smRNA-FISH

Table of Contents

List of Tables

List of Figures

List of Abbreviations, Acronyms and Symbols

5caC – 5-carboxylcytosine

5fC – 5-formylcytosine

5hmC – 5-hydroximethylcytosine

5mC – 5-methylcytosine

Amot – Angiomotin

AVE – Anterior Visceral Endoderm

BMP – Bone Morphogenic Protein

Cdx2 – Caudal Type Homeobox 2

C – Cytosine

CpG – 5'—C—phosphate—G—3'

DMSO – Dimethyl Sulfoxide

DNMTs – DNA Methyltransferases

DVE – Distal Visceral Endoderm

EPI – Epiblast

E – Embryonic day

ESCs – Embryonic Stem Cells

FACS – Fluorescence-Activated Cell Sorting

FBS – Foetal Bovine Serum

FGF – Fibroblast Growth Factor

Fgfr – Fibroblast growth factor receptor

H3K4me3 – trimethylation of lysine 4 in H3 histone

H3K9ac – acetylation of lysine 9 in H3 histone

H3K9me3 – trimethylation of lysine 9 in H3 histone

H3K14ac – acetylation of lysine 14 in H3 histone

H3K27me3 – trimethylation of lysine 27 in H3 histone

H3K36me3 – trimethylation of lysine 36 in H3 histone

hESCs – human Embryonic Stem Cells

HMTs – Histone Methyltransferases

ICM – Inner Cell Mass

LIF – Leukaemia Inhibitor Factor

mESCs – mouse Embryonic Stem Cells

mRNA – messenger Ribonucleic Acid

Oct4 – Octamer-binding transcription factor 4

PBS - Phosphate-buffered saline

PE – Primitive Endoderm

PRC2 – Polycomb Repressive Complex 2

PS – Primitive Streak

RNA-seq – RNA sequencing

RT-qPCR - Reverse Transcription quantitative Polymerase Chain Reaction

smRNA-FISH – single molecule RNA Fluorescent In Situ Hybridization

Sox2 – Sex determining region Y-box 2

TDG – Thymine-DNA Glycosylase

TE – Trophectoderm

TET – Ten Eleven Translocation

TFs – Transcription Factors

TGFβ – Transforming Growth Factor β

1. Introduction

1.1 Early Mammalian Development

After fertilization, the zygote undergoes a series of divisions into smaller blastomeres, creating a morula. While the first division is meridional, the second is rotational, with one of the blastomeres dividing meridionally and the other equatorially (Figure 1.1; reviewed in Zernicka-Goetz, 2005).

In the zygote, the genome is not being transcribed, and the molecules necessary to the first division were inherited from maternal messenger RNA (mRNA), deposited in the oocyte during its formation. In different mammals, zygotic genome activation happens at different timepoints: 2-cell stage in mice (*Mus musculus*), 4-cell stage in pigs (*Sus scrofa domesticus*), 8-cell stage in sheeps (*Ovis aries*), and between the 4- and 8-cell states in humans (*Homo sapiens sapiens*; Piko and Clegg, 1982; Crosby, Gandolfi and Moor, 1988; Braude, Bolton and Moore, 1988). Due to the experiments being performed mainly on mouse, I shall focus on this model system from now on.

In mouse embryos, after division from 4- to 8-cell stage, blastomeres endure a process of compaction. This process consists of cell shape changes with previously round and separated blastomeres becoming a compact and spherical group of cells. It is accompanied by formation of tight and gap junctions, in a calcium dependent manner, and accumulation of the adhesion molecule E-cadherin at cell-cell junctions (Ducibella and Anderson, 1975; Hyafil, Baninet and Jacob, 1981; Peyriéras et al., 1983).

With compaction, each blastomere becomes polarized, allowing its division in different orientations (Ziomek and Johnson, 1980). If a blastomere divides symmetrically (perpendicularly to the apical-basal axis), it gives rise to similar daughter cells belonging to the outer layer of the morula. If it divides asymmetrically (along the apical-basal axis), it gives rise to different daughter cells, with the more apical belonging to the outer layer and the more basal becoming completely internalised. Two cell populations thus arise in the embryo (Johnson and Zyomek, 1981): while cells in the outer layer will later develop into the trophectoderm (TE), the inner cells will contribute to the inner cell mass (ICM) of the blastocyst. This is the first cell fate decision happening in the developing mouse embryo (Fleming, 1987).

1.1.1 The First Cell Fate Decision

Inner cells of the morula will give rise to the ICM and the outer cells will become TE, although with some plasticity: if moved, cells can still convert to the other fate, depending on their position (Hillman, Sherman and Graham, 1972; Handyside, 1978; Rossant and Lis, 1979; Spindle, 1978). Polarization of the blastomeres during compaction and following asymmetric divisions made the inner and outer cells distinct due the presence of different components. In the case of inner cells, these cells express transcription factors (TFs) of the pluripotency network, mainly Oct4 (octamer-binding transcription factor 4) and Sox2 (sex determining region Y-box 2; Palmieri et al., 1994; Nichols et al., 1998; Avilion et al., 2003). Oct4 and Sox2 from maternal origin were already present in the zygote, but zygotic expression only starts at morula stage. In the blastocyst, expression of these TFs will be later restricted to ICM (Palmieri et al., 1994; Avilion et al., 2003).

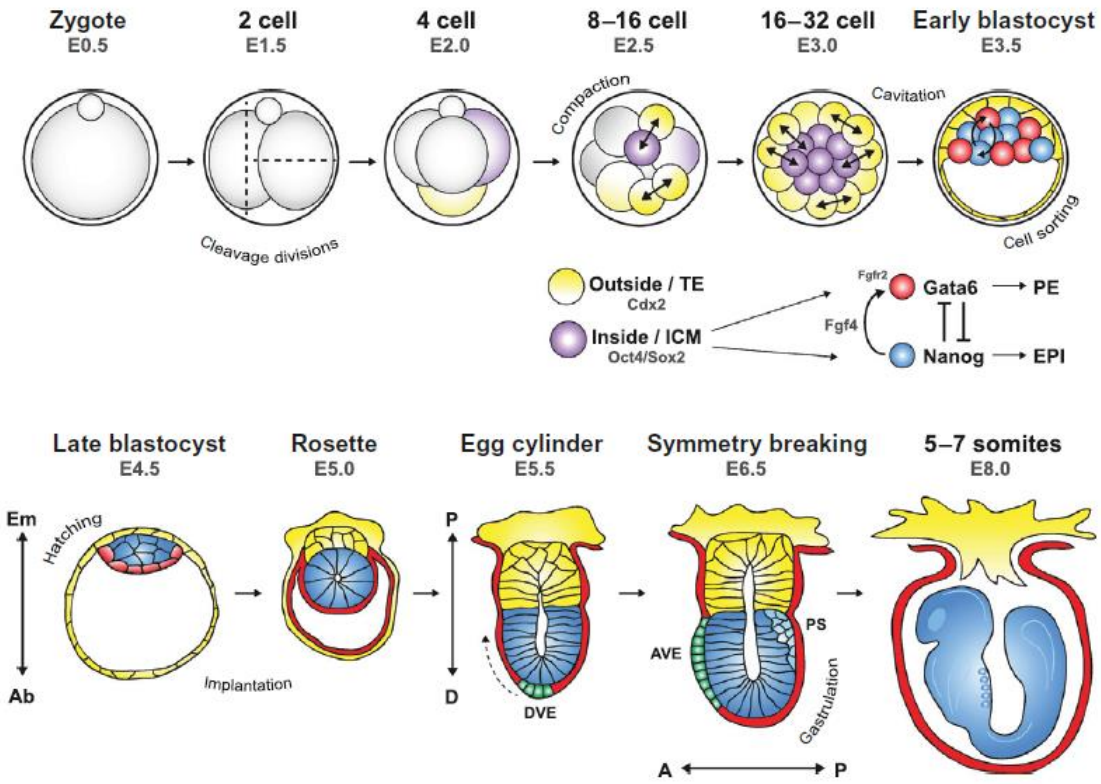

Figure 1.1 – Early Mouse Embryonic Development. Through series of cleavage divisions, the zygote gives rise to the morula at E3.0 (embryonic day 3) with two already separated cell populations. Cells on the inside will give rise to the ICM and outside cells will give rise to the TE of the blastocyst. The localization of the ICM determines the embryonic-abembryonic axis of the embryo, that later turns into proximal-distal axis, following implantation. As the blastocyst matures, ICM cells will be specified into epiblast (EPI) or primitive endoderm (PE), due to Nanog, Gata6, Fgf4 and Fgfr2 expression and activity. Once the blastocyst has matured, it hatches from the zona pellucida, implants and continues to develop. The EPI lineage reorganizes into a polarized rosette with the opening of a lumen. The antero-posterior axis will be determined by the migration of the distal visceral endoderm cells (DVE) to form the anterior visceral endoderm (AVE) to one side of the egg cylinder. The primitive streak (PS) formation on the posterior side of the embryo, brakes the symmetry of the embryo and determines the beginning of gastrulation (Graham and Zernicka-Goetz, 2016).

Cells at the outer region of the morula express different TFs, like Cdx2 (caudal type homeobox 2) that is one of the main TFs involved in TE differentiation and specifically expressed in the outer cells of the morula (Niwa et al., 2005; Strumpf et al., 2005). *Cdx2* expression is regulated by Tead4, which is uniformly expressed in all cells, independently of their position (Nishioka et al., 2008). Although uniformly expressed, Tead4 only activates *Cdx2* expression in outer cells due to the differential activity of its transcriptional activators Yap and Taz (Nishioka et al., 2009). In inner cells, phosphorylation of Yap and Taz by the Hippo pathway kinases Lats1/2 prevent their migration to the nucleus to act with Tead4 in the activation of *Cdx2* expression (Nishiola et al., 2009).

Although all components of the Hippo pathway necessary to phosphorylate Yap and Taz being present in all cells, the critical factor is angiomotin (Amot), which is located at adherent junctions in inner cells and in the apical region in outer cells. Amot apical position prevents its interaction with Lats1/2, not phosphorylating Yap/Taz, not sequestering it in the cytoplasm, being able to migrate to the nucleus and activate *Cdx2* expression together with Tead4 (Figure 1.2; Hirate et al., 2013; Leung and Zernicka-Goetz, 2013).

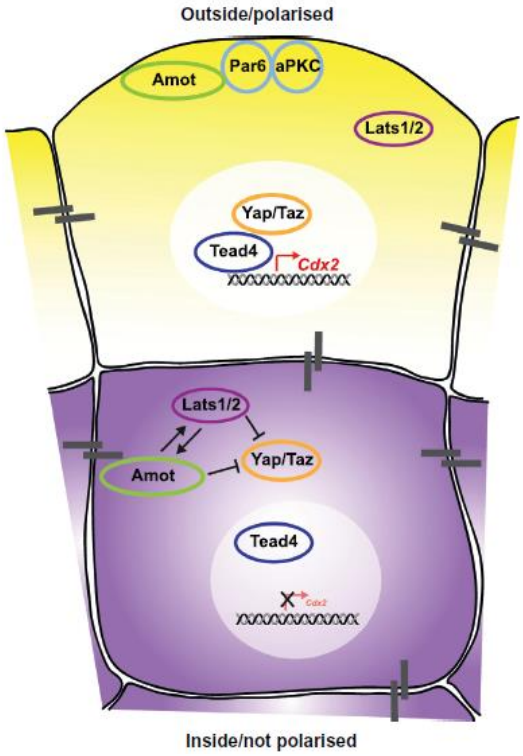

Figure 1.2 – Polarity in the First Cell Fate Decision. Cdx2 expression is dependent of Tead4 and Yap/Taz activity. In outside cells (yellow) Yap/Taz are not phosphorylated by Lats1/2 due to the apical positioning of Amot, being able to migrate to the nucleus and coactivate Cdx2 expression. In inner cells of the morula (purple), Lats1/2 can phosphorylate Yap/Taz, sequestering it in the cytoplasm and inhibiting Cdx2 expression (Graham and Zernicka-Goetz, 2016).

In the beginning of morula, *Cdx2* and *Oct4* are co-expressed in the all cells. Sorting occurs through a process of cross-regulation, in which each TF activates their own expression and represses expression of the other. During this Oct4/Cdx2 competition, cell fate is still plastic, but it will evolve to complete exclusion of each TF to individual cells, which acquire a specific fate ($Oct4^+$ cells become ICM and $Cdx2^+$ become TE; Niwa et al., 2005).

During the initial differentiation of TE and ICM, cells are still in direct contact, but through a process of cavitation, the blastocoel is formed. Blastocoel is a cavity inside the blastocyst, filled with fluid secreted by trophoblast cells. Trophoblast cells express a Na^+/K^+-ATPase that pumps Na^+ to the cavity of the forming blastocyst, followed by the osmotic transport of water (Cross, 1973; Watson and Kidder, 1988). With increased volume of the blastocoel, the ICM sits in one side of the blastocyst, with some cells contacting the trophoblast and the others the blastocoel fluid.

At this point, there is a surge of cell death (Handyside and Hunter, 1986). The reason is not yet completely understood, but one of the main explanations might be cell-cell competition, with less fit cells being eliminated. In mESCs, it has been shown that cells with defective signalling or defective gene expression are eliminated through competition (Sancho et al., 2013).

1.1.2 The Second Cell Fate Decision

When the blastocyst forms at E3.5, ICM cells are already heterogeneous in expression of some TFs. Two of the most heterogeneously expressed TFs are Nanog and Gata6. During EPI/PE specification, PE

progenitor cells will migrate and form a monolayer separating the EPI cells from the blastocoel. The migration depends on the expression of LamininB1 and Dab2, cellular adhesion molecules, whose expression is induced by Gata6 (Figure 1.3; Chazaud et al., 2006).

Although specification of ICM into EPI or PE only occurs with maturation of the blastocyst, it started during the internalization and asymmetric divisions at the morula stage due to differential fibroblast growth factor (FGF) signalling (Yamanaka et al., 2010). The internalization of blastomeres occurs in two distinct rounds of asymmetric cell divisions (from 8- to 16-cell and 16- to 32-cell stages; Chazaud et al., 2006). Blastomeres internalised in the first round of cell divisions are biased towards EPI while blastomeres internalised in the second round of divisions are biased towards PE (Morris et al., 2010). At the 16-cell stage, after the first round of internalisation, Fgf4 (fibroblast growth factor-4) is upregulated in inner cells by Nanog, inducing the expression of its receptor, Fgfr2 (fibroblast growth factor receptor 2), in the outer cells of the morula. Thus, in the second round of internalization, the internalised blastomeres present higher expression of Fgfr2, being able of higher response to FGF signalling. As Fgf4 produced by the first internalised cells reaches its receptor in the later internalised cells, FGF signalling cascade induces the expression of Gata6 in these cells. As the first internalised cells receive less FGF signalling, Gata6 is not activated. Then, by a process of mutual repression, Nanog and Fgf4 will be restricted to the future EPI cells and Fgfr2 and Gata6 to the future PE cell (reviewed in Lanner and Rossant, 2010).

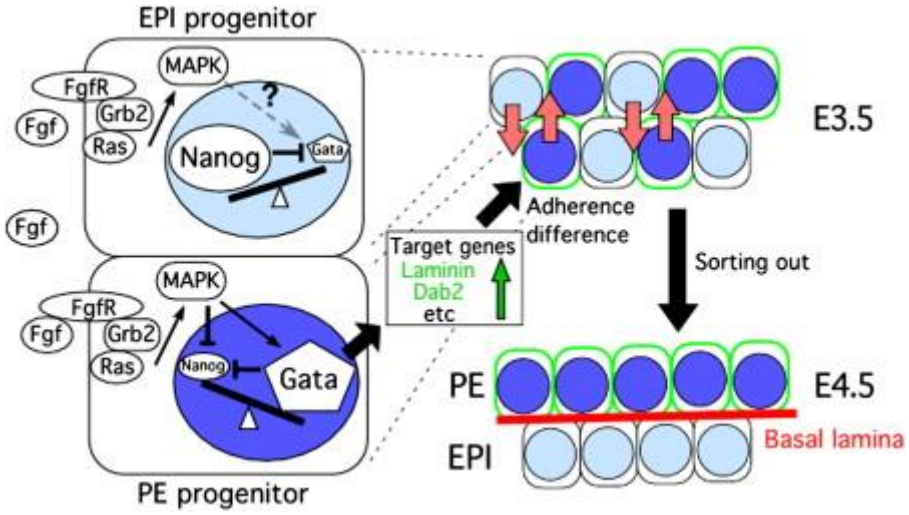

Figure 1.3 – FGF Signalling in ICM Lineage Commitment. Fgf4 is produce in the first internalised cells (Nanog expressing cells), inducing the expression of its receptor, Fgfr2, in the later internalised cells. In these last cells, FGF signalling cascade induces the expression of Gata6, that will repress Nanog expression and induce the expression of LamininB1 and Dab2, adhesion molecules responsible for the migration and sorting out of PE. Then, by mutual exclusion, Nanog and Fgf4 will be restricted to EPI cells and Gata6 and Fgfr2 restricted to PE cells (Chazaud et al., 2006).

Besides formation of EPI and PE, another event that must occur after maturation of blastocyst is the escape from the zona pellucida. This is a layer of extracellular matrix protecting the blastocyst since its formation in the oocyte. This is a critical process without which blastocysts cannot implant into the uterus and continue the development of the embryo (Perona and Wassarman, 1986).

1.2 Capturing Pluripotency

The zygote is a totipotent cell, being able to differentiate into cells with capacity to develop into all embryonic and extraembryonic structures of an embryo. As they divide, the initial blastomeres remain totipotent, only starting to be specified into different lineages with the formation of the morula. Once the blastocyst is formed, two lineages are defined and only the cells from the ICM remain pluripotent. ICM cells have the capacity to differentiate into all embryonic tissues, but lost the capacity to differentiate into extraembryonic tissues, that are restricted to TE. Throughout embryonic development, as cells differentiate, they lose some of their potential until they reach the final state, a completely differentiated cell. However, some progenitor multipotent cells remain in the adult organism (reviewed in Wagers and Weissman, 2004).

Pluripotency is a temporary state in embryonic development, starting around the early blastocyst stage (E3.5), with the emergence of epiblast precursors, and persisting in their descendants until being completely dismantled prior to somitogenesis, at E8.0 (Osorno et al., 2012).

The limited number of cells in a blastocyst and its rapid development make it very difficult to study the differentiation potential of epiblast cells. As such, embryonic stem cells (ESCs) were established as an alternative to study this process.

1.2.1 mESCs as a Study Model

Mouse ESCs (mESCs) are cultured cells derived from the ICM of a blastocyst (Evans and Kaufman, 1981; Martin, 1981), maintaining similar characteristics to the cells from which they were derived. These cells maintain the pluripotency and self-renewal properties characteristic of the ICM/EPI cells of blastocysts, being able to be expanded without losing their potential and, given the right cues, differentiate into all cell types (Keller, 1995).

Due to these characteristics, when injected into a blastocyst at an appropriate stage according to the culture conditions, they contribute to the formation of all lineages of the embryo, including the germ line (Beddington and Robertson, 1989; Huang et al., 2012).

1.2.2 Maintaining the Pluripotency

Initially, mESCs were cultured over a layer of mitotically inactivated fibroblast feeder cells, in media previously conditioned by embryonal carcinoma cells, containing also calf serum or foetal/new-born calf serum (Evans and Kaufman, 1981; Martin, 1981). These culture conditions are now known to contain signalling molecules involved in LIF, BMP and Wnt signalling, the main signalling pathways involved in the maintenance of pluripotency and self-renewal (Figure 1.4).

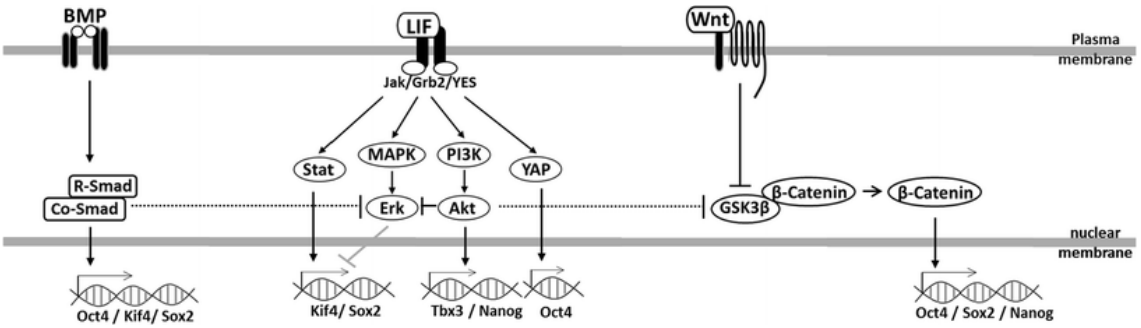

Figure 1.4 – **The Major Signalling Pathways Regulating Pluripotency in mESCs.** BMP, LIF and Wnt signalling pathways work together to activate the pluripotency network in ESCs (Chen et al., 2016).

LIF Signalling

Leukaemia inhibitor factor (LIF) is a member of the interleukin 6 family of cytokines that is secreted by murine embryonic fibroblasts and stimulates self-renewal while repressing differentiation of mESCs (Smith et al., 1988; Williams et al., 1988).

LIF binds to a heterodimer of LIF receptor beta and gp130 and can activate the signalling pathways of Jak-Stat3, PI3K-Akt, YES-YAP and MAPK-Erk. While the three first pathways stimulate pluripotency, the last stimulates differentiation (Boeuf et al., 1997; Niwa et al., 1998; Armstrong et al., 2006; Niwa et al., 2009).

BMP Signalling

Bone morphogenic proteins (BMPs) are members of the transforming growth factor β (TGFβ) family of cytokines. BMPs bind to a heterodimer of type I and type II receptor serine/threonine kinases. Receptor II phosphorylates receptor I, activating it, leading to phosphorylation of Smad proteins to regulate transcription of target genes (reviewed in Shi and Massagué, 2003).

Wnt Signalling

Wnt signalling works through inhibition of the glycogen synthase kinase 3 β (GSK3β). Binding of Wnt to its receptor phosphorylates GSK3β, leading to the release of β-catenin, which migrates to the nucleus and activates expression of genes of the pluripotency network, mainly Oct4 (Kelly et al., 2011).

Nowadays, mESCs are cultured in more defined media that allow for the maintenance of sub-states within the pluripotency continuum. There are two most common culture media. One depends on the use of foetal bovine serum or serum substitutes, together with LIF (hence named Serum/LIF), while the other uses two specific inhibitors: PD0325901, a MEK inhibitor (upstream of Erk) and CHIR99021, a GSK3β inhibitor. This culture condition is known as 2i (Ying et al., 2008).

mESCs have a transcriptionally hyperactive genome and this property is considered to be one of the hallmarks of pluripotency (Efroni et al., 2008). The transcription dynamics create heterogeneity that can be modulated by different culture conditions.

mESCs grown in presence of 2i have higher and more homogeneous expression level of pluripotency genes and a low expression of lineage-affiliated genes (Marks et al., 2012). These cells are in a naïve state and are considered to resemble pre-implantation epiblasts of E4.5 embryos (Boroviak et al., 2014),

although some studies showing that they might resemble cells from embryos between E1.5 (2-cell stage) and E3.5, with embryonic and extraembryonic potential (Macfarlan et al., 2012; Gonzalez et al., 2016).

Recent studies have found that prolonged maintenance of mESCs in 2i impairs their developmental potential, one of the main characteristics of ESCs. The use of the inhibitors induces a widespread DNA methylation loss, which can only be reverted in mESCs cultured for short periods of time, when transplanted into blastocyst. Once cultured for longer periods in 2i, mESCs lose the capacity to methylate imprinted genomic regions that are developmentally essential, resulting in developmental arrest (Choi et al., 2017; Yagi et al., 2017).

When mESCs are cultured in Serum/LIF, expression of pluripotency genes is more heterogeneous, and expression of lineage-affiliated genes is higher (Efroni et al., 2008). These mESCs are considered to be similar to post-implantation epiblasts of E5.5 embryos (Boroviak et al., 2014), although some studies placing them closer to E4.5 embryos (Gonzalez et al., 2016). Due to higher heterogeneity in this culture condition, several sub-populations can be delimited and are thought to correspond to different embryonic stages, spanning from E3.5 to E5.5 (Papatsenko et al., 2015).

These culture conditions maintain mESCs in a pluripotency state with similarities to the embryo, but they also need to stabilize it to be able to expand indefinitely, something that does not happen in the embryo. During *in vitro* derivation of mESCs from the ICM, there are many genes with differential expression, as cells replace their genetic program of differentiation into a program of maintenance of self-renewal and pluripotency. These modifications in gene expression are also accompanied by changes in epigenetic modifications and in the expression of microRNAs (Tang et al., 2010).

1.2.3 Pluripotency Network

The pluripotency network maintains the special characteristics of ESCs, with the core of this network being composed by the TFs Oct4, Sox2 and Nanog (Figure 1.5).

Oct4

Oct4, encoded by the *POU5f1* (POU domain, class 5, transcription factor 1) gene, is a member of the mammalian POU family of transcription factors, whose expression is necessary for the formation of the pluripotent population in an embryo (Nichols et al., 1998). Although necessary, the expression level needs to be tightly regulated as it has been shown that either increasing or decreasing its expression causes loss of pluripotency. Overexpression causes differentiation into primitive endoderm and mesoderm while decrease of expression causes dedifferentiation into trophectoderm (Niwa et al., 2000).

Sox2

Sox2 is a member of the Sry-related HMG (high-mobility-group) box family of transcription factors and is known to bind to the same regions of the genome as Oct4, acting individually and synergistically. The Oct4-Sox2 joint function is possible by interaction of their POU and HMG domains (Chew et al., 2005). Like Oct4, Sox2 is also necessary for the formation of the pluripotent cells of a blastocyst, as its deletion has been shown to cause cells to go into a trophectodermal fate (Masui et al., 2007).

Nanog

Nanog is a homeodomain protein that forms functional dimers through its tryptophan-rich domain (Mullin et al., 2008; Wang et al., 2008), with dimerization being essential to the interaction with other pluripotency factors (Wang et al., 2006).

In the embryo, *Nanog* expression is restricted, being expressed only in the ICM, EPI and germline progenitor cells (Chambers et al., 2003; Mitsui et al., 2003; Yamaguchi et al., 2005). Due to its critical role in epiblast determination, epiblast cells fail to develop in *Nanog* mutants, with cells shifting into an extraembryonic endoderm lineage (Mitsui et al., 2003).

Although being necessary for cells to reach the pluripotent state, pluripotent cells can be maintained and expanded in the permanent absence of Nanog. These cells are, however, prone to differentiation (Chambers et al., 2007). On the other hand, overexpression of Nanog confers a higher capacity of self-renewal, independently of the presence of cytokines like LIF (Chambers et al., 2003).

Besides being very important for developmental progression of the mouse embryo, Nanog is not one of the essential TFs to reprogram fibroblasts into induced pluripotent stem cells (Takahashi and Yamanaka, 2006), like Oct4 and Sox2. Klf4 is the necessary factor and it was later found to have a role in the activation of Nanog (Zhang et al., 2010), the necessary factor to establish and maintain pluripotency (Figure 1.5; Silva et al., 2009).

TFs of the pluripotency network have two main functions: they work together to maintain the pluripotency network active, while at the same time repressing the expression of genes involved in differentiation pathways (Marson et al., 2008). The Oct4-Sox2 complex has been shown to bind to *Nanog* promoter (van den Berg et al., 2008), besides their own promoters, regulating its expression (Catena et al., 2004; Chew et al., 2005). Nanog does not interact physically with Oct4 and Sox2, but shares many of their binding sites, possibly belonging to the same activation or repressive complex (Chen et al., 2008; Marson et al., 2008).

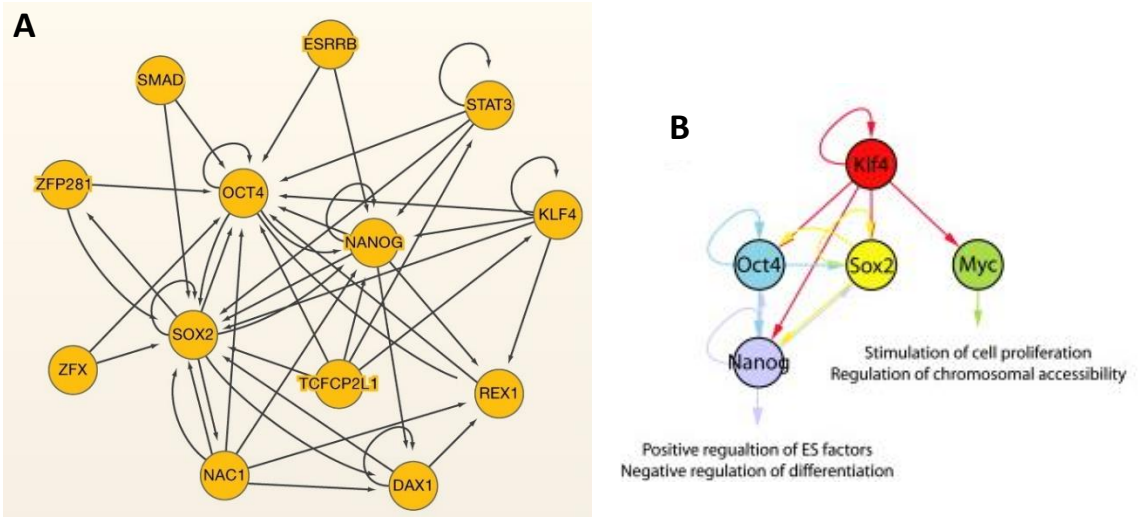

Figure 1.5 – The Pluripotency Network. (**A**) The main elements of the pluripotency network in ESCs, with Oct4, Sox2 and Nanog having connections with the majority of the others. (**B**) Transcriptional network of necessary factors to induce pluripotency in differentiated cells (Klf4, Oct4, Sox2 and Myc) and their relationship to Nanog. In both cases, arrows indicate the direction of transcriptional regulation (adapted from Orkin and Hochedlinger, 2011; Kim et al., 2008).

1.3 Heterogeneity: Creating Possibilities

In embryos, gene expression heterogeneity is present from the moment of compaction, being also detected in ESCs. In mESCs, some pluripotency-associated genes exhibit very heterogeneous expression, including *Nanog* (Figure 1.6; Chambers et al., 2007; Singh et al., 2007), *Rex1* (Toyooka et al., 2008), *Stella* (Hayashi et al., 2009) and *Pecam1* (Furusawa et al., 2004).

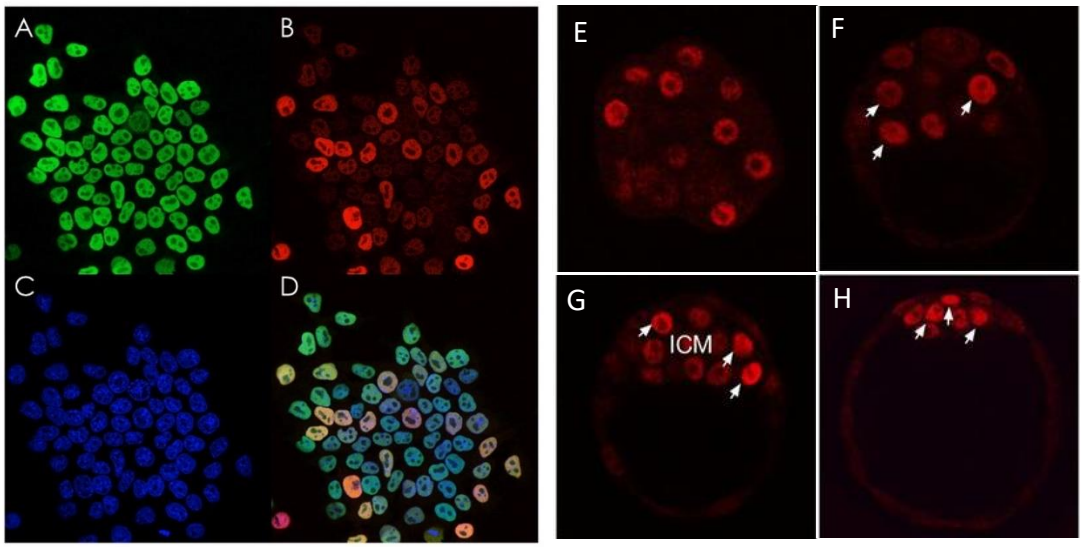

Figure 1.6 – Nanog Expression in mESCs and Embryos. Immunofluorescence staining for: (A) Oct4, (B) Nanog, (C) staining with DAPI and (D) an overlay of A-C. Oct4 staining is relatively homogeneous, whereas Nanog expression levels vary widely. (E-H) Nanog expression in the developing embryo, from late morula to blastocyst (E3.0 - E4.0; adapted from Roeder and Radtke, 2009; Acampora, Giovannantonio and Simeone, 2013).

This heterogeneity might be one of the ways the blastocyst, and therefore mESCs, have found to be able to differentiate into various differentiation paths. ESCs with lower expression of pluripotency TFs might be in a more differentiable state and be first to differentiate when conditions are met. Thus, heterogeneity might create a window of opportunity.

Gene expression can occur as a constitutive process, characterized by a Poissonian-like distribution of transcripts, or episodic processes characterized by stochastic activation (ON) and inactivation (OFF) periods of gene expression, resulting in discontinuous production of mRNA. The episodic process of gene expression is termed transcriptional bursting and is responsible for creating heterogeneity. Transcriptional bursting is influenced by several factors, including the chromatin environment dependent on histone modifications and nucleosome occupancy, transcription factors availability and DNA looping (reviewed in Nicolas, Philips and Naef, 2017). These factors modulate burst size and frequency, generating heterogeneity on gene expression. In humans, transcriptional bursting has been shown to be the predominant mode of gene expression (Dar et al., 2012).

In mESCs, Oct4 and Sox2 are expressed at high and homogeneous levels, whereas Nanog expression levels vary widely, with cells spanning from high expression to very low, or even no expression. mESCs not expressing Nanog, or expressing low levels, remain pluripotent, as shown by their expression of Oct4 (Figure 1.6; Chambers et al., 2007).

Nanog is necessary for the establishment of pluripotency in ICM cells, but once established, it can be maintained in its absence, as shown by experiments in mESCs (Chambers et al., 2007). This suggests that the pluripotency network can operate with different levels of Nanog.

It is therefore possible to define the existence of pluripotent cells expressing high or low levels of Nanog, hereafter designated as High-Nanog and Low-Nanog, respectively. Although remaining pluripotent, Low-Nanog mESCs are prone to differentiation, have a lower capacity to self-renew and present higher expression of differentiation-associated genes, than High-Nanog mESCs (Chambers et al., 2007; Abranches et al., 2013; Abranches et al., 2014).

Expression of differentiation-associated genes in pluripotent cells has been termed as lineage-priming, occurring mainly in Low-Nanog, a permissive state. These cells are expressing different lineage specific genes, each trying to instate their differentiation lineage and induce differentiation of the pluripotent ESCs (Martinez-Arias and Brickman, 2011). Lineage-priming does not imply commitment, as primed cells can revert to a naïve pluripotency state.

High-Nanog cells are in a pristine state of pluripotency, irresponsive to differentiation, where the pluripotency network is fully active and repressing the expression of lineage-affiliated genes. On the contrary, Low-Nanog cells are in a primed state of pluripotency, exploring the possible differentiation options, while remaining pluripotent. Hence, the Low-Nanog state might correspond to an initial stage of differentiation (Figure 1.7).

Nanog heterogeneity has been attributed to stochastic fluctuations in gene expression in individual mESCs. These fluctuations might create a window of opportunity, where mESCs can move between different metastable cell states, accompanied by fluctuations in gene expression. These metastable states are in a dynamic continuum of multiple interconvertible states, in which ESCs have different responsiveness to differentiation stimuli (Graf and Stadtfeld, 2008).

Fluctuations in Nanog expression occur independently of the culture conditions, being an inherent characteristic of mESC, although the Low-Nanog population of cells in the pluripotency ground-state (2i) being smaller, when compared with Serum/LIF (Abranches et al., 2014).

As these states are functionally and molecularly distinct, Nanog heterogeneity can confer an advantage to the mESCs population, as it maintains cells in naïve (High-Nanog) and primed (Low-Nanog) states of pluripotency. Thus, the population can explore multiple differentiation options while maintaining a pool of naïve pluripotent cells (Abranches et al., 2014).

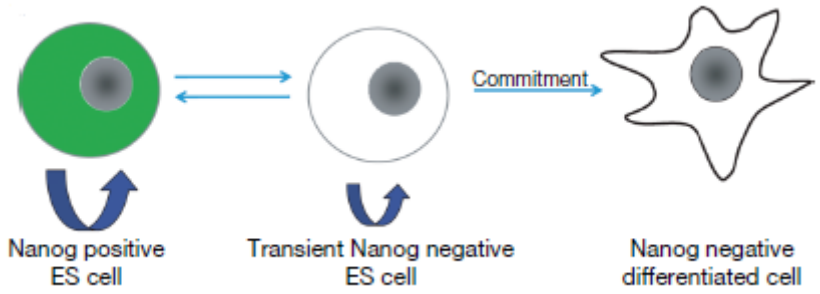

Figure 1.7 – Central Role of Nanog in Differentiation. mESCs fluctuate widely between pluripotent states with different levels of Nanog. Nanog positive ESCs are in a naïve state of pluripotency while Nanog negative ESCs are in a pluripotent primed state. These transient cells can return to the naïve state by re-expressing Nanog or can commit to differentiation (adapted from Chambers et al., 2007).

1.4 Epigenetics in Development

Besides the role of TFs, epigenetic modifications also play a central role in the maintenance of pluripotency and in cell fate specification (reviewed in Surani, Hayashi and Hakjova, 2007). Here, the focus will be on DNA methylation and chromatin modifications.

1.4.1 DNA methylation

Throughout development, the CpG (5'—C—phosphate—G—3') methylation pattern changes dynamically to modulate the expression of specific genes necessary to form different structures of the embryo (Figure 1.8).

The gametes are methylated at different proportions, with the oocyte being more hypomethylated than sperm. After fertilization, both genomes face global demethylation in the egg, with sperm genome starting to be demethylated right after the formation of the zygote, while the oocyte genome is only demethylated during the initial cleavages (Smith et al., 2012). The only exceptions to this global demethylation process are the imprinted control regions inherited from the progenitors, as their methylation is maintained during this phase of demethylation and reprograming (Bartolomei and Ferguson-Smith, 2011).

This genome wide demethylation occurs until blastocyst formation, as methylation reaches its lower level in the ICM (Smith et al. 2012). Following this permissive state, where most of the genome is demethylated, it starts to be remethylated as cells differentiate into the three germ layers, with the exception of primordial germ cells, that will be completely demethylated and remethylated to form the gametes (Smith et al 2012).

In concordance with their developmental resemblance with ICM cells, mESCs cultured in 2i present an hypomethylated genome, with similar methylation levels to cells from early blastocysts. On the other hand, mESC cultured in Serum/LIF exhibit a hypermethylated genome, with methylation levels closer to the post-implantation epiblast (Habibi et al., 2013). This methylation pattern corroborates the gene expression experiments that postulate that 2i and Serum/LIF correspond to different stages of the blastocyst.

CpG methylation is catalysed by DNA methyltransferases (DNMTs). The members of this enzyme family with methyltransferase activity are Dnmt1, Dnmt3a and Dnmt3b, all working with Dnmt3l as an accessory protein. Dnmt1 is involved in methylation maintenance after DNA replication, and Dnmt3a/b are essential for *de novo* methylation in both ESCs and early mouse development. These enzymes catalyse the methylation at the 5-carbon of the cytosine (C) residue within the cytosine-guanine dinucleotides (CpG), resulting in the formation of 5-methylcytosine (5mC; Hermann et al., 2004; Okano et al., 1999).

Once methylated, Ten-Eleven Translocation (TET) enzymes, including Tet1, Tet2 and Tet3, can demethylate 5mC. TETs are 2-oxoglutarate (2OG)- and Fe(II)-dependent enzymes. Besides the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), Tet enzymes can also convert 5hmC into 5-formylcytosine (5fC) and 5fC into 5-carboxylcytosine (5caC; Ito et al., 2011). Once completely oxidised by TETs, 5caC is specifically recognized and excised by thymine-DNA glycosylase (TDG) (Figure 1.9; He et al., 2011).

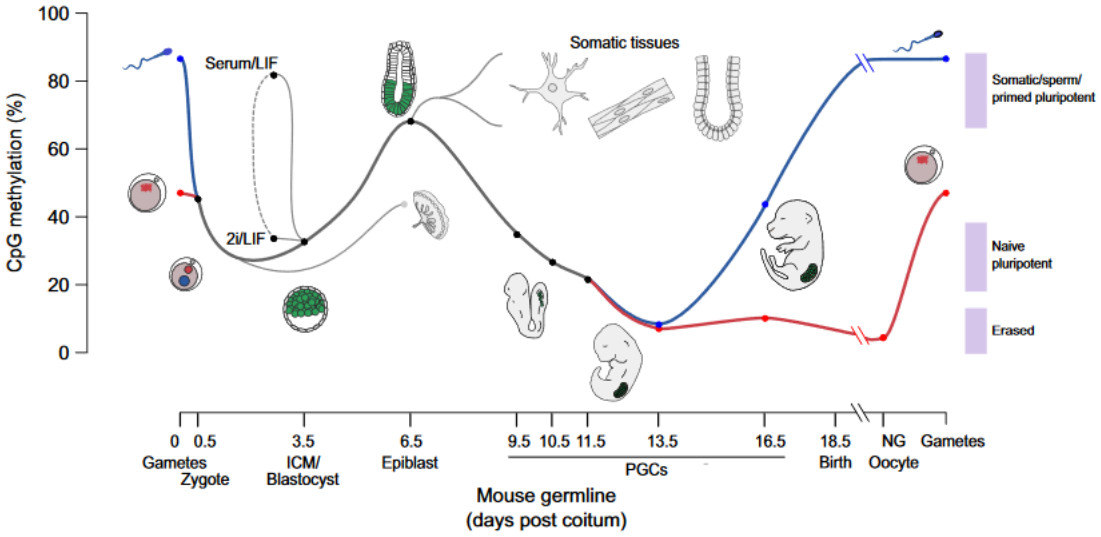

Figure 1.8 – Global CpG Methylation Levels Throughout Development. Paternal (blue line) and maternal (red line) genomes lose methylation until the blastocyst stage (E3.5), regaining it as cell fate specification and differentiation occurs. In primordial germ cells, the methylation is completely erased between E6.5 and E13.5, as these cells emerge from the epiblast. In these cells, methylation level will be regained in a sex-specific manner, as they are specified. ESCs can be cultured from the ICM, and based on the culture conditions remain hypomethylated in 2i or become hypermethylated in Serum/LIF. These methylation states are interconvertible by changing culture medium (dashed line; Lee et al., 2014).

TET enzymes have different cellular distributions and roles, with Tet1 being mainly expressed in ESCs and in the embryo. It has a very important role, regulating Nanog expression, and when depleted, causes downregulation of Nanog and upregulation of TE and PE specific genes (Ito et al., 2010). While Tet1 is predominantly found in the promoter regions and transcription start sites of its target genes, Tet2 is mainly associated with gene bodies of highly expressed genes in mESCs (Huang et al., 2014). Unlike Tet1 and Tet2, Tet3 main role in early development is the epigenetic reprogramming of the zygotic paternal DNA following fertilization, and in the activation of zygotic Oct4 expression (Gu et al., 2011).

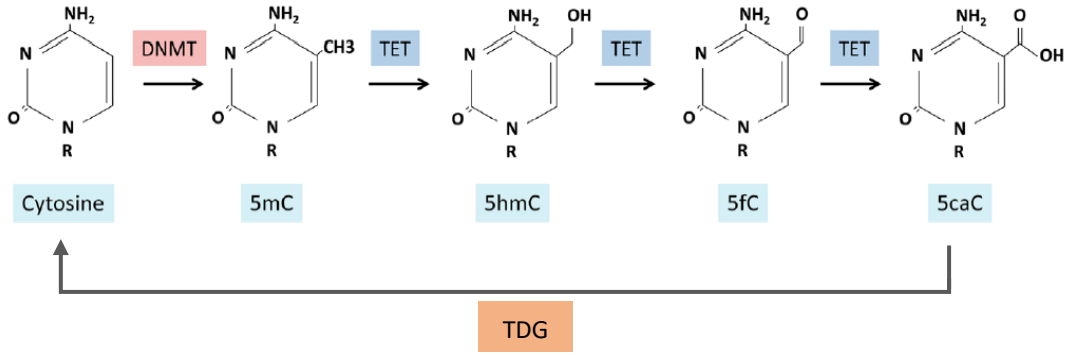

Figure 1.9 – Cytosine Methylation and Demethylation Cycle. Cytosines are methylated at the 5-carbon of the cytosine by DNMTs, resulting in 5mC. 5mC is then oxidized by TETs through a series of reactions, until 5caC, which will be excised by TDG to restore the initial unmodified cytosine (adapted from Nightingale, 2006).

Besides a role in transcriptional activation, through CpG demethylation, Tet1 can also contribute to the repression of genes, mediating the recruitment of PRC2 to CpG-rich promoters, enriched in 5hmC. In Tet1 mutants, 5mC in maintained and recruitment of PRC2 to these promoters is impaired (Wu et al., 2011).

1.4.2 Chromatin Modifications

Besides the roles of DNA methylation in transcriptional regulation of gene expression, chromatin modifications are also important regulators of gene expression.

The chromatin basic unity is the nucleosome. It is composed of 147 base pairs of DNA wrapped around an octamer of histones, composed of two of each histone H2A, H2B, H3 and H4 (Luger et al., 1997). The nucleosome is initially formed by standard histones, which can later have their N-terminal tails modified. The possible modifications are acetylation of lysine residues, methylation of lysines and arginines, phosphorylation of serines and threonines, ubiquitylation of lysines, sumoylation of lysines, ADP ribosylation of glutamic acid, and isomerization of prolines, which are catalysed by different enzymes/complexes (reviewed in Kouzarides 2007).

These modifications affect the accessibility of the DNA to other proteins (non-histones), increasing or decreasing it. Trimethylation of lysine 4 and acetylation of lysines 9 and 14 in H3 histone (H3K4me3, H3K9ac and H3K14ac, respectively) are characteristic of active promoters (Bernstein et al., 2002; Liang et al., 2004) and H3 lysine 36 trimethylation (H3K36me3) is enriched throughout bodies of transcribed genes (Bannister et al., 2005). While these modifications are present in regions of more open chromatin, trimethylation of lysines 9 and 27 of H3 histone (H3K9me3 and H3K27me3, respectively) are associated with more compact and silenced chromatin (Lachner et al., 2001; Cao et al., 2002).

ESCs are enriched in H3K4me3, H3K9ac, H3K14ac, H3K36me2, and H3K36me3, consistent with an open chromatin conformation and a high level of transcription (Efroni et al., 2008). During differentiation, open chromatin and its characteristic modifications are substituted by closed chromatin and repressive modifications, as H3K9 methylation (Wen et al., 2009)

Another characteristic of ESCs is the overlap of opposite chromatin modifications, H3K4me3 and H3K27me3, in the same genes (Figure 1.10). This very specific occurrence is known as bivalent chromatin and is mainly found in genes encoding developmental regulators, characterized by low expression levels in both mESCs and hESCs (human ESCs; Azuara et al., 2006; Bernstein et al., 2006; Pan et al., 2007; Zhao et al., 2007). This bivalent state maintains these genes in a poised state, being easily activated or repressed when needed. During differentiation into specific lineages, some bivalent genes will increase expression following loss of the repressive modification, H3K27me3, and maintenance of the positive modification, H3K4me3. If a bivalent marked gene is not necessary to the chosen differentiation path, the positive modification H3K4me3 is lost, being followed by enrichment in H3K9me3 (Figure 1.10; Bernstein et al., 2006; Meshorer et al., 2006).

The repressive modification, H3K27me3, is catalysed by the polycomb repressive complex 2 (PRC2; Cao et al., 2002). PRC2 is a chromatin-modifying complex comprising four main components: Ezh1/2, Suz12, Eed and RbAp46/48 (also known as RBBP7/4). Ezh2, a histone-lysine N-methyltransferase, is responsible for the methylation of H3K27 (reviewed in Margueron and Reinberg, 2011).

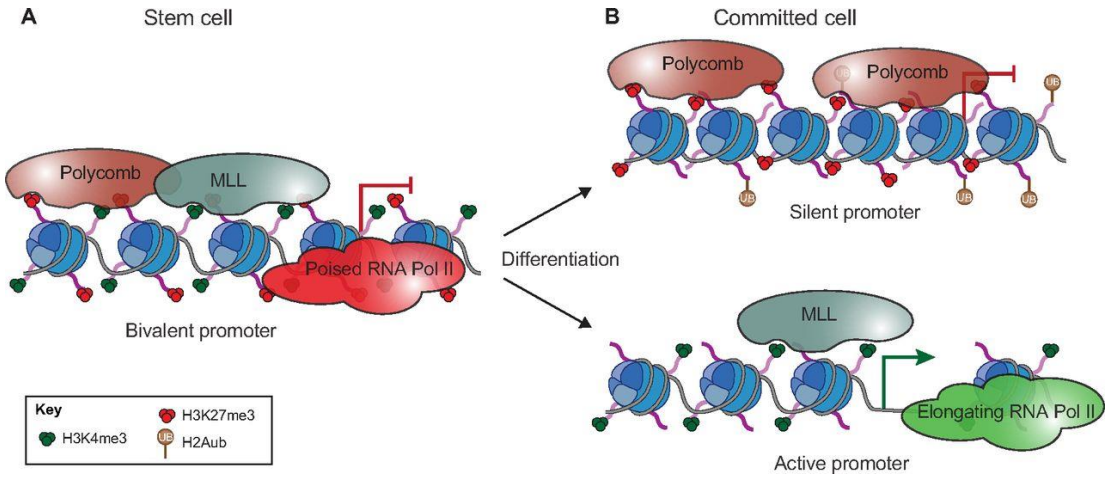

Figure 1.10 – Role of Bivalent Chromatin in Developmentally Relevant Genes. (**A**) Bivalent promoters are characterized by the existence of repressive (H3K27me3) and activation (H3K4me3) marks, deposited by Polycomb and MLL complexes, respectively. During differentiation (**B**), if the gene is not necessary, it is silenced, losing H3K4me3 and gaining other repressive marks. In necessary genes, H3K27me3 is lost, H3K4me3 is increased, and the previously poised RNA Polymerase II transcribes the gene (Aloia et al., 2013).

After implantation, the epiblast needs to be maintained in an undifferentiated pluripotent state. The maintenance of the pluripotent state and repression of differentiation is dependent of PRC2 activity, as it has been shown that depletion of Eed, Ezh2 and Suz12 result in lethality in early post-implantation stages, with embryos displaying severe developmental and proliferative defects (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004).

The activation mark H3K4me3 is catalysed by MLL-family histone methyltransferases (HMTs), which include 4 different proteins in mammals (MLL1-4), together with three structural components RbBP5, Ash2L and WDR5. Mll1 and Mll2 are the methyltransferases essential for embryonic development (Dou et al., 2006; Denissov et al., 2014). In ESCs, WDR5 is regulated by Nanog and Oct4, being necessary for the maintenance of pluripotency, inducing differentiation and reducing self-renewal when depleted (Ang et al., 2011).

1.5 Nanog as a Key Regulator of Priming Gene Expression

Pluripotent mESCs are highly heterogeneous in the expression of some TFs. Based on Nanog heterogeneous expression, two separated populations of cells can be defined, Low-Nanog and High-Nanog. In High-Nanog cells, the pluripotency network is fully active, maintaining cells in a naïve state of pluripotency, irresponsive to differentiation signals. The opposite happens in the Low-Nanog state, which remains pluripotent due to expression of Oct4 and Sox2, but also expresses lineage-affiliated (or priming) genes (Chambers et al., 2007; Abranches et al., 2013; Abranches et al., 2014).

This observation led to the hypothesis that Nanog might be a regulator of these priming genes. Thus, in the High-Nanog state, Nanog should be actively repressing the expression of the priming genes, whereas in the Low-Nanog state its absence allows their expression (Abranches et al., 2014).

Being a transcription factor, Nanog binds to its targets and recruits the transcription machinery. Nanog has been shown to interact with Tet1, recruiting it to the promoters of pluripotency associated genes like Oct4, during reprogramming. In mESCs, Nanog and Tet1 bind to the promoters of genes associated with

maintenance of pluripotency and lineage commitment, with Nanog stabilizing Tet1 binding (Costa et al., 2013).

The genes associated with lineage commitment are also associated with bivalent chromatin, being repressed by histone methylation through PRC2. PRC2 modifies histones in specific regions and has previously been shown to bind 5hmC-enriched promoters (Wu et al., 2011).

Therefore, we hypothesize that Nanog might repress the expression of priming genes through recruitment of Tet1, which converts 5mC into 5hmC. PRC2 recognizes these 5hmC-rich regions and catalyses the deposition of H3K27me3, ensuring the repression of these genes (Figure 1.11).

Previous laboratory work led to the identification, by RNA-Seq, of genes upregulated and downregulated in Low- and High-Nanog states (unpublished data). A chromatin enrichment analysis, performed in Enrichr (Chen et al., 2013; Kuleshov et al., 2016), showed that the genes upregulated in the High-Nanog state are associated with proteins involved in the pluripotency network, whereas priming genes (upregulated in the Low-Nanog state) are mainly associated with members of PRC2 (unpublished data).

Some of the priming genes detected by RNA-seq have already been analysed by similar methods to those here employed, and the preliminary results showed that some might be regulated by our proposed model (like *Sox3*), while others do not (*Car2*; unpublished results).

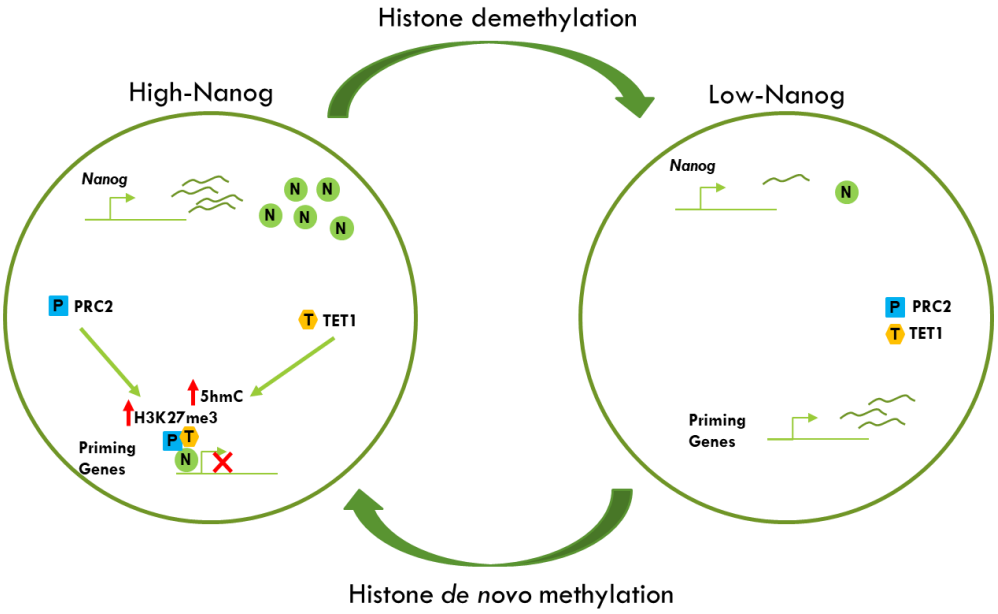

Figure 1.11 – Nanog Role in Priming Genes Regulation. mESCs can be in High-Nanog or Low-Nanog states. In High-Nanog state, Nanog is repressing the expression of priming genes through Tet1 and PRC2. In Low-Nanog mESCs, the absence of Nanog allows the expression of priming genes and mESCs can explore the possible differentiation pathways. The opposite should happen in High-Nanog mESCs, where Nanog recruits Tet1 to the regulatory regions of priming genes; where Tet1 will catalyse the conversion of 5mC into 5hmC. PRC2 will recognize these hypomethylated regions and trimethylate H3K27, ensuring the repression of the priming genes. Transition from High-Nanog to Low-Nanog implies H3K27 demethylation around these genes, whereas the opposite transition requires methylation.

2. Aims

Since the discovery of ESCs, abundant research has been done to understand the mechanisms of how these cells reach and maintain their pluripotent state. Besides knowledge-driven research, much of it was motivated by all the potential medical applications. Stem cells can differentiate into every somatic cell found in an adult organism, opening doors to stem cell-based therapies in regenerative medicine.

ESCs are derived in culture from pluripotent cells of blastocyst epiblasts, although presenting some epigenetic modifications that allow them to stay in this state, indefinitely. It is not yet clear if the heterogeneity and fluctuations of Nanog expression have a functional role in lineage priming in mESCs and mouse embryonic development. So far, most of the experiments were performed in populations of mESCs, masking heterogeneity and lineage-priming. It is therefore essential to study these processes at the single cell level. Until now, several techniques have been developed to study gene expression in single cells, including reverse transcription quantitative polymerase chain reaction (RT-qPCR), RNA-sequencing (RNA-seq) and single molecule RNA fluorescent in situ hybridization (smRNA-FISH), among others, with smRNA-FISH being the only highly quantitative method to study gene expression (reviewed in Kanter and Kalisky, 2015).

Here, we use smRNA-FISH, which allows the quantification of the exact number of mRNA molecules existent in a cell (Raj et al., 2008), to study the role of Nanog heterogeneity and dynamics on priming gene expression.

To test our hypothesis of priming gene regulation by Nanog, via recruitment of Tet1 and PRC2, small molecule modulators of Tet1 and PRC2 activities will be used. PRC2 activity can be inhibited by the GSK343 compound (Verma et al., 2012), while Tet1 activity can be stimulated with ascorbic acid (AA, also known as Vitamin C; Blaschke et al., 2013; Yin et al, 2013). The effects of these treatments were assessed by quantification of the exact number of various mRNA transcripts in single cells, through smRNA-FISH.

This project aims to:

1. Describe pluripotency and priming gene transcriptional heterogeneity at the single cell level, through quantification of mRNA transcripts in Serum/LIF conditions;
2. Demonstrate the higher frequency of lineage-priming in the Low-Nanog stage, correlating *Nanog* and priming-gene expression;
3. Assess the role of Nanog in the regulation of priming genes, through Tet1 and PRC2.

3. Materials and Methods

3.1. Materials

3.1.1 Cell Lines

Two mESC lines, E14tg2a (from now on called E14) and Nd (from Nanog dynamics), were used. Nd is a Nanog:VNP reporter cell line derived from E14, containing a BAC transgene with a short-lived fluorescent protein (VNP – Venus NLS Pest) under Nanog regulation, previously developed at the lab (Abranches et al., 2013).

3.1.2 Reagents

The reagents, solutions, smRNA-FISH probes and detection filters used throughout this project are listed in the following tables.

Table 3.1 – List of Reagents Used Throughout the Experiments. The reagents are listed with information relative to suppliers, catalogue numbers and stock/working solutions. RT- room temperature

Reagent	Supplier	Cat. Number	Stock	Working Stock
2-mercaptoethanol	Sigma	M-7522	RT	0.1 M in H_2O, 4ºC
Catalase	Sigma	C-3515	-	4ºC
DAPI (4',6-diamidino-2-phenylindole)	Sigma	D9542	1 mg/mL in PBS, -20ºC	1.5 µg/mL in PBS, 4ºC
Dextran Sulphate	Sigma	D8906	-	4ºC
DMSO (Dimethyl sulfoxide)	Sigma	D-2650	RT	RT
Dow Corning High Vacuum Grease	Sigma	Z273554	-	RT
Foetal Bovine Serum	Hyclone	SH30070	-20ºC	Heat-inactivated -20ºC
Formaldehyde 37%	Sigma	252549	-	RT
Formamide	Ambion	AM-9342	-	4ºC
Gelatin 2%	Sigma	G-1393	4ºC	0.1% in PBS, 4ºC
Glucose	Sigma	C-6152	-	RT
Glucose oxidase	Sigma	G2133	37 mg/mL in 50 mM Sodium Acetate, -20ºC	3.7 mg/mL in 50 mM Sodium Acetate, -20ºC
Glutamine	GIBCO	25030-123	200 mM	100x, -20ºC

Table 3.1 – continuation.

Reagent	Supplier	Cat. Number	Stock	Working Stock
GMEM	GIBCO	21710-025	-	1x, 4ºC
GSK343	Sigma	SML0766	1mM in DMSO, -80ºC	1mM in DMSO, 4ºC
Ascorbic Acid (AA; L-Ascorbic Acid2-phosphate)	Sigma	A8960	50mg/mL in sterile H_2O, 4ºC	50mg/mL in sterile H_2O, 4ºC
Non-essential aminoacids	GIBCO	11140-035	-	100x, 4ºC
PenStrep	GIBCO	15140-122	-	100x, -20ºC
Propidium Iodide	Invitrogen	P-3566	1 mg/mL, 4ºC	1 ng/mL, 4ºC
Sodium Pyruvate	GIBCO	11360-039	-	100x, -20ºC
Triton	Sigma	T8787	-	RT
Trypan blue	Sigma	T8154	RT	0.4% in PBS, RT
Trypsin	GIBCO	25090-028	2.5% (v/v), -20ºC	0.25% in PBS, 4ºC

Table 3.2 – List of Solutions Used Throughout the Experiments. The solutions used are listed with their components and storage conditions. RT – room temperature

Solutions	Components	Storage
Anti-fade buffer	100 uL Glox buffer buffer; 1uL catalase; 1uL 3.7mg/mL glucose oxidade (prepare fresh)	RT
EtOH 70%	35mL 95% EtOH; 15 mL miliQ H2O	RT
Gelatin 0.1%	2.5mL 2% gelatin; PBS up to 50mL	4ºC
Glox buffer	850uL H_2O; 100uL 20x SSC; 40uL 10% glucose; 10uL Tris 1M PH8; 10uL 10%Triton (prepare fresh)	RT
Glucose 10%	1.5g glucose; H_2O up to 15mL	RT
GMEM 1x	80% (v/v) GMEM; 1% (v/v) Glutamine; 1% (v/v) Pen-Strep; 1% (v/v) Sodium Pyruvate; 1% (v/v) Non-essential Aminoacids; 10% (v/v) FBS; 0,001% (v/v) of 2-mercaptoethanol	4ºC
Hybridization buffer	1g dextran sulfate; 7mL miliQ H_2O; 1mL formamide; 1mL 20x SSC; miliQ H_2O up to 10mL (store at -20ºC)	-20ºC
SSC 20x	87.65g 3M NaCl; 44.11g 0.3M Sodium citrate; miliQ H_2O up to 500mL	RT
TE	10mM Tris; 1mM EDTA pH=8;	RT
Triton 10%	1mL Triton X-100; PBS up to 10mL	RT
Trypsin 0.01%	5mL 0.25% Trypsin; PBS up to 50mL	4ºC
Wash buffer	5mL 20x SSC; 5mL formamide; 500uL 10% Triton; 40mL miliQ H_2O	RT

Table 3.3 – List of Probes Used for mRNA Detection in smRNA-FISH experiments. Probe list for each target gene and their correspondent fluorophores. For each target gene, 25 to 45 unique oligonucleotides (20 nucleotides) probes were previously designed and dissolved in TE, at a concentration of 1 to 12μM. Each probe is labelled with a fluorophore: alexa594, cy5 or tmr.

Gene	Oligo Sequences
Car2 (alexa595)	gtgacaggcagaggtgacag, aggggaggagaccgtggag, tgattggggcagagcagaag, ctccattggcaatggggaag, tgctgtgtcaatgtccacag, gatatgagcagaggctgtag, gttgacaatgctcttggacg, tcaacgttaaaggagtggcc, attgtcctgagagtcatcaa, atctgtaggagtcactgagg, ccccagtgaaagtgaaactg, gttccagtgaaccaagtgaa, caaaacagccaatccatccg, gaagttagcaaaggccgcac, caggaagaagggagcaagga, tatgtccagtagtccaagtt, acgatccaggtcacacattc, ctcgctgctgacagtaatgg, cattgaagttcagcgtacgg, ctattctttagcggctgagc, cttaaaggacgctttgatct, ttagctacagagaggcggtc, caaatcacccagcctaactg, acaataccagatgcgagtcg, agcacaacggatgagaggta, gtctcatgatgtggacttgt, tttgcctaagttacttcagt, atccattgtgttgtggtatg
Crabp2 (tmr)	ccagagctctaggcttttat, ccagagctctaggcttttat, cctttgcagaacagatcctt, ggctaaagatactttgctgt, ttcaactagaacactggacg, gcaaggtggctttctcttta, ttgccagaaaagttaggcat, agttttccgatcggatgatc, tctcctgtttgatctcgact, aagttaatctccgtggttcg, caccaaactcttacagggtc, attggtcagttctcggctcc, ttgtcaggatcagctctcca, tagaccctggtgcacacaac, cgtaggcactcactctcgga, gaagtcgtctcaggcagttc, tttgtgtcctgtagcgggca, agtaaccctcaggagtaagg, gagggttctgagcttctgcc, ttccaggaaggaccctggtg, tagaggccagagagacagtc, aatcacacagactacaagga
Dnmt3b (tmr)	Attcagatgtctgctgtctc, ataatgcactcctcataccc, ggtcactgaagttcccatta, taattcagaaggctggagac, atcatctctgtctccatctc, agccattcccatcatctact, ggtgagctttggcattagaa, ttgctgaagatgatgctcga, acttcttccatgaagtcgac, aagtcaactgatggggtact, tatccataccctcctgatct, atactctgtgctgtctccat, ggtcacctattccaaactct, aggagaagcccttgatcttt, aacttgccatcaccaaacca, tttgtcagcagagatctcag, ttaaagtgctggctgaacag, cagcttattgaaggtagcca, gtacatggccttcctataag, tgcttcttgttgggtttgag, gcaccttcgacttattaacc, aagttcctactgtctgaacg, ttcgacttttgttctcgcgt, aagcagcagagtcattggtt, tatttgtcttgaggcgcttg, ttggtgacttcagaagccat, aaacagcggtcttccagatt, acagggttcttctttccaca, tgtagaagagctctaggaag, tgatagccgtcctcatcata, atctgcagcagcttgtgtta, tgttccaatctttcctgcgt, tagtgaagaagtcttgcagg, aattcttccaggtcaggatc, aattgctgggtacaacttgg, actctaatgggcctcctttt, ccgttgcaattccatcaaac, aactccttgagcaccaagta, gaggcaatgtacttttccac, ttaacagttcccacagcgat, ccggacgtcattgacatatt, ggccccactcttcaatattt, agagatcattgcatgggctt, atataaacctttgcgggcag, aaactcgaagaagagccttc, gggtataattcagcaagtgg, atccagaagaatggacggtt, ttcatggccacaacattctc
Fgf5 (alexa594, tmr)	caagcgctgtggatcaggtg, cttcgggagtgagacgcttc, gagtctcccgggttcctagg, aagaaaacgtcgcgctactt, ctatgttccgagccgcttcc, cgaagggctccactggaaac, ccactctgcagtacaggctg, tagatctgcagatggaaacc, agccattgactttgccatcc, cttaacacactggcttcgtg, ttcctacaatcccctgagac, ttgttgctgaaaactcctcg, agtcatccgtaaatttggca, tcttggaatctctccctgaa, gcggacgcataggtattata, cgtaccactctcggcctgtc, gagacgtgttggggtttgac, cttgaacctgggtaggaagt, gtgaaggaaagttccggttg, cacctttggtttcaccggtg, tcagtctgtacttcactggg, gagcatcatccaaagcgaaa, tgacgcctgtatagagagtt, gctgtatccgagtttccttc, aaatgacctgacttccacac, tagttccagtcaaagcgaaa, cccacttctgtttcgactta, ctccctggatcgctacagag, atcctcggtgatctgaagag, cctcaaatcacgaagccttt, gaaaacgcagaatgctaacc, gttccaggctatcaaactgg
Nanog (alexa594, cy5, tmr)	aaatcagcctatctgaaggc, cagaaagagcaagacaccaa, gaagtcagaaggaagtgagc, actcagtgtctagaaggaaa, ggttttaggcaacaaccaaa, cgagggaagggatttctgaa, cacactcatgtcagtgtgat, cagaactaggcaaactgtgg, ttcccagaattcgatgcttc, aaaaactgcaggcattgatg, agcaagaatagttctcggga, cagagcatctcagtagcaga, gaagaggcaggtcttcagag, tgggactggtagaagaatca, tcaggacttgagagcttttg, cttgttctcctcctcctcag, gagaacacagtccgcatctt, ctgtccttgagtgcacacag, tgaggtacttctgcttctga, gagagttcttgcatctgctg, atagctcaggttcagaatgg, gaaaccaggtcttaacctgc, ttgcacttcatcctttggtt, tcaaccactggtttttctgc, ttctgaatcagaccattgct, gatactccactggtgctgag, ggatagctgcaatggatgct, cagatgcgttcaccagatag, aagttgggttggtccaagtc, gtctggttgttccaagttgg, aaagtcctccccgaagttat, ctgcaactgtacgtaaggct, caaatcactggcagagaagt, tagtggcttccaaattcacc, ctaaaatgcgcatggctttc, ataattccaaggcttgtggg, tggagtcacagagtagttca, agatgttgcgtaagtctcat, gctttgccctgactttaagc, tttggaagaaggaaggaacc, caaatcactggcagagaagt, tagtggcttccaaattcacc, ctaaaatgcgcatggctttc, ataattccaaggcttgtggg, tggagtcacagagtagttca, agatgttgcgtaagtctcat, gctttgccctgactttaagc, tttggaagaaggaaggaacc

Table 3.3 – continuation.

Gene	Oligo Sequences
Sox2 (alexa594)	ccgtctccatcatgttatac, tccgggctgttcttctggtt, ataccatgaaggcgttcatg, ttctcctgggccatcttacg, atctccgagttgtgcatctt, tcggacaaaagtttccactc, ttataatccgggtgctcctt, tcatgagcgtcttggttttc, ggaagcgtgtacttatcctt, tagctgtccatgcgctggtt, ttgctccagccgttcatgtg, tcctgcatcatgctgtagct, tgcatcggttgcatctgtgc, tcatggagttgtactgcagg, ttcatgtaggtctgcgagct, agtaggacatgctgtaggtg, ttgaccacagagcccatgga, tgggaggaagaggtaaccac, aggtacatgctgatcatgtc, tgggccatgtgcagtctact, agtgtgccgttaatggccgt, aaaatctctccccttctcca, cccaattcccttgtatctct, tactctcctctttttgcacc, ctgcggagatttttttcct, tttttccgcagctgtcgttt, aatttggatgggattggtgg, tagtcggcatcacggttttt, gaagtcccaagatctctcat, ctgtacaaaaatagtccccc, tatacatggtccgattcccc, gcgtagtttttttcctccag, cctaacgtaccactagaact, aagacttttgcgaactccct, ccggagtctagctctaaata, ctgtacaaaagttgcttgca, gattgccatgtttatctcga, caagaaccctttcctcgaaa, aagctgcagaatcaaaaccc, ccttgtttgtaacggtccta, ccagtacttgctctcatgtt, aacaagaccacgaaaacggt, acaatctagaacgtttgcct, gatatcaacctgcatggaca, gggtaggattgaacaaaagc, cggaaaataaaaggggggaa, ccaataacagagccgaatct, tatacatggattctcggcag
Sox3 (tmr)	Ttctctcgagctggtcgcat, cgggcttctctcacctgatg, acaccatgaacgcgttcatg, ggttctccagggccatcttg, atctcggagttgtgcatctt, catcggtcagcagtttccag, cttggcctcgtcgatgaacg, gtacttgtagtccgggtact, gagcagcgtcttggtcttgc, gcagcgagtacttgtccttc, ttcacgtgcgtgtacgtgtc, ctcctgcacgagcgagtagg, atgtcgtagcggtgcatctg, gagagctgggctccgacttc, ggtacatgctgatcatgtcg, accgttccattgaccgcagt, gagcaaagctaaacagcaag, catcttcggtacaaggcaac, gacagttacggccaaacttt, ggacttctcgcttttgtaca, gctctagcaagtcccatttc, gaacctaggaatccgggaag, gacattttcaactgcaacag, gggcaacctcactcagttct, tggaggcattgcagttcttg, aacattggctttagctgtcg, aactcaacagcctaaacgcg, agcaaatagatcactgcaga, gaacgaaatgcgtacacgaa, actttgaaaaaacctggaac

Table 3.4 – Optical Filters for mRNA Detection in smRNA-FISH Experiments. The optical filters used are listed, along with their characteristics, including excitation, beam splitter and emission wavelengths

Filter	Excitation [nm]	Beam splitter or dichroic [nm]	Emission [nm]	Supplier
TMR	539/21	556	576/31	Chroma
Alexa594	590/10	610	630/30	Omega
Cy5	640/30	660	700/75	Chroma
DAPI	365/12	395	> 397	Zeiss

3.2. Methods

3.2.1 Expansion of Embryonic Stem Cells

All mESCs were manipulated in a sterile laminar flow hood class II, type A/B3 to maintain sterility and prevent contaminations.

For mESCs expansion, cells were cultured in GMEM 1x supplemented with 10% foetal bovine serum (FBS) and 2ng/mL leukaemia inhibitory factor (LIF), Serum/LIF conditions, on gelatine-coated (0.1% (v/v)) dishes at 37ºC in a 5% (v/v) CO_2 incubator.

Cell morphology was assessed daily on a brightfield microscope (Leica DM IL) and mESCs were passed every other day, and plated at a constant density of $3x10^4$ cell/cm^2.

mESCs from the frozen stock ($3x10^6$ cells/vial), at -80ºC in liquid nitrogen, were thawed in a 37ºC water bath. During thawing, the partially frozen cells were ressuspended in pre-heated Serum/LIF medium. To remove the residues of dimethyl sulfoxide (DMSO) from the medium, mESCs were centrifuged at 1200 rpm (rotations per minute) for 4 min (minutes), the medium aspirated and mESCs ressuspended again in pre-heated medium. mESCs were then plated on gelatine-coated dishes. Medium was changed after 6 hours to ensure the complete removal of DMSO. After 24 to 48h (hours), depending on plate confluence, cells were passed.

For every passage, mESCs were washed twice in phosphate-buffered saline (PBS) and dissociated with trypsin for 2 min, at 37ºC. After dissociation, to neutralize the enzymatic reaction of trypsin, mESCs were immediately ressuspended in Serum/LIF, centrifuged at 1200 rpm for 4 min, supernatant removed and mESCs ressuspended in medium. mESCs density and viability were determined by counting in a haemocytometer, using the trypan blue dye exclusion method to exclude unviable cells. mESCs were then plated at the appropriate density.

E14 and Nd mESCs were always cultured in parallel to allow monitoring of Nanog:VNP levels.

3.2.2 Chemical Modulators Assay

In the chemical modulators assays, mESCs were cultured as previously described for 3 passages to allow their stabilization. mESCs were then grown in Serum/LIF supplemented with the chemical modulators, ascorbic acid (1mg/mL), GSK343 (1 μM) and both. As a control, mESCs were cultured in Serum/LIF supplemented with DMSO (0.1%).

This assay had the duration of 48 hours. To ensure the even distribution and maintenance of the modulators/DMSO, the medium was changed at 24h. At assay ending, Nanog:VNP levels were assessed and mESCs fixed for smRNA-FISH.

3.2.3 Flow Cytometry

In order to quantify the percentages of Nanog positive cells at every passage, approximately $5x10^5$ mESCs were taken from cell culture, ressuspended in PBS and Nanog:VNP levels assessed in a BD Accuri C6 flow cytometer. Live cells were gated based on forward and side scatter parameters and non-viable cells were excluded by propidium iodide dye. As a control, non-fluorescent E14 mESCs were used.

3.2.4 Fluorescence-Activated Cell Sorting (FACS)

mESCs were expanded as previously described, dissociated, resuspended in PBS and transferred to FACS tubes coated in 1% BSA in PBS. mESCs were sorted based on Nanog:VNP fluorescence by a BD FACSARIA III into pure Low-Nanog:VNP and High-Nanog:VNP populations. After sorting, cells were ressuspended in Serum/LIF and plated as described for the chemical modulators assay, or fixed for smRNA-FISH.

3.2.5 Single Molecule RNA Fluorescent *in Situ* Hibridization (smRNA-FISH)

Single molecule RNA Fluorescent in Situ Hybridization (smRNA-FISH) is a technique that allows for the labelling and posterior quantification of single mRNA molecules in single cells (Raj et al., 2008). This method is designed to give the specificity of the single cell analysis and, at the same time, the whole population dynamics of several mRNAs.

Following cell culture and dissociation, mESCs were washed twice in PBS and fixed in formaldehyde for 10min at room temperature. After fixation, mESCs were washed twice in PBS to remove formaldehyde residues and ressuspended in 70% ethanol. Once in ethanol and stored at 4ºC, these cells do not suffer mRNA degradation.

An appropriate amount of fixed cells was then used for each smRNA-FISH experiment. Cells were centrifuged at 3000rpm for 2min, ethanol was removed and cells were washed in 850µL of wash buffer, followed by cell resuspension in the hybridization mix and overnight incubation at 37ºC. The hybridization mix was prepared fresh from 100µL of hybridization buffer and 1µL of each smRNA-FISH probe.

In the next day, cells were washed with 850µL of wash buffer and incubated for 30min at 37ºC to remove the non-hybridized probes. Next, cells were washed again and incubated with wash buffer and with 1µL of DAPI (to stain cell nucleus) for 30min at 37ºC. Finally, cells were washed in Glox buffer solution (prepared fresh) and resuspended in 10µL of anti-fade buffer (prepared from Glox buffer, catalase and glucose oxidase). Cells were then kept at 4ºC until mounting, between 2 and 4 hours.

Before imaging, 5µL of cell suspension in antifade were mounted between slide and coverslip. In order to decrease cell height and increase signal quality, cells were slightly smashed by applying pressure on the coverslip with tweezers. At last, the sample was sealed with Dow Corning High Vacuum Grease to prevent liquid evaporation.

Following this mounting step, cells were imaged on a motorized inverted widefield fluorescence microscope, Zeiss Cell Observer, equipped with a 100x (1.4NA) oil-immersion objective, Zeiss Axiocam 506 camera, large cage incubator and the appropriate filter sets to the fluorophores used. Microscope incubation was set to 24ºC to prevent thermal drift of the sample and loss of cell focus.

For each experiment, 100 to 120 images, comprising a Z-stack of 20 optical sections with an interval of 0.3µm, were acquired.

3.2.6 Data Analysis

Following imaging, images were converted from Zeiss image format (.czi) to most common tiff image format (.tiff), without loss of quality, by custom MATLAB software developed at Instituto de Medicina Molecular, Lisboa.

Images were analysed in a MATLAB custom software developed at Raj Laboratory for systems biology, in University of Pennsylvania (Raj et al., 2008; Raj and Tyagi, 2010; available at https://bitbucket.org/arjunrajlaboratory/rajlabimagetools/wiki/Home). In the software, each cell was manually delimited and the number of spots (transcripts) accounted for each of the channels used.

Counts of mRNAs per cell were extracted from MATLAB and analysed in RStudio software. Analysis consisted on the calculation of the statistical measures associated with the counts of mRNAs: average, median, standard deviation, variance, minimum, maximum, number of cells analysed, Fano factor and coefficient of variation. Correlations between the different mRNAs in a single cell were also calculated (spearman correlation) and analysed. Besides these statistical values, the graphical environment of RStudio provided a great tool for the visualization of mRNA counts distribution histograms and correlation dot plots, using the graphical package ggplot2 (Wickham, 2011).

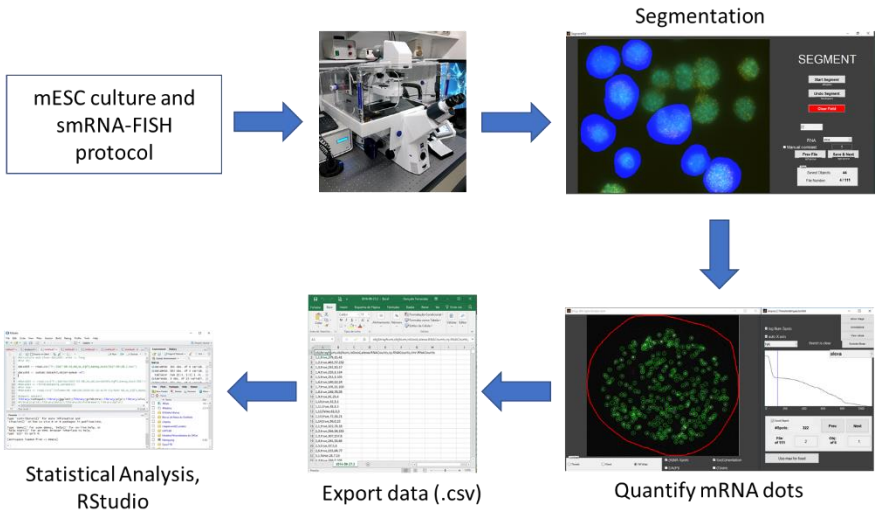

Figure 2.1 – smRNA-FISH Workflow. Workflow for smRNA-FISH experimental procedures and data analysis.

4. Results

In order to detect and characterize transcriptional heterogeneity among mESCs, smRNA-FISH was employed. smRNA-FISH allows the detection and quantification of single mRNAs in individual cells. Using fluorescent probes that cover most of the mRNA, one can detect individual mRNAs in fixed cells as individual fluorescent dots.

As smRNA-FISH allows the parallel quantification of several genes in the same cell, its transcriptional state might be inferred by analysing the correlations between the expression levels of different genes.

For each experiment, histograms representing the frequency of cells expressing counts of mRNAs, were plotted and analysed. Transcript distribution shape is informative of the transcription dynamics of the gene. In Gaussian-like distributions, most cells express an average number of transcripts, and few cells express less or more transcripts, indicating continuous active transcription. Long tailed distributions, where most cells contain few transcripts and few cells express high levels of transcripts, is an indication of bursting transcription (Raj and van Oudenaarden, 2009).

Besides visual examination of histograms, Fano factor (FF) and coefficient of variation (CV) were also calculated. FF is defined as the ratio between variance to mean, and CV is the ratio between standard deviation to mean. In Poisson distributions, there is low cell-to-cell variation and FF is 1 and CV is 0, while in long tailed distributions, transcription bursts create high cell-to-cell variation, resulting in higher FFs and CVs (Raj and van Oudenaarden, 2009).

Due to the non-normality of gene expression distributions, Spearman correlations (Spearman, 1904) were calculated for each experiment. The correlation coefficient varies between -1 and 1, corresponding to negative and positive correlations, respectively. Correlations closer to 1 indicate that cells exhibiting high expression of one gene are likely to co-express the other, while correlations closer to -1 are an indication of cells exhibiting high expression of one gene and low expression of the other ("anti-correlation").

To characterize mRNA distributions in mESC populations, statistical analysis included calculation of mean, median, standard deviation (SD), Variance, FF, CV, minimum, maximum and number of cells. The corresponding results are presented as histograms of mRNA distributions, dispersion plots with the correlations between two or three genes, and tables with statistical measures for each gene.

4.1 Morphology of mESCs and Nanog:VNP Dynamics

mESC were cultured in Serum/LIF and morphology was assessed daily in an inverted bright field microscope. mESCs grew in dome shaped clusters, with the cells on the edge of the clusters presenting irregular forms, with elongations, a morphology characteristic of differentiating cells. No morphological differences were found between the two mESC lines used.

Use of the Nd mESC cell line allowed the monitorization of Nanog:VNP levels. This cell line was previously developed and validated in the laboratory, with VNP mimicking *Nanog* expression (Abranches et al., 2013). Nanog:VNP levels were assessed at every passage and were always within the normal range previously obtained for mESCs cultivated in Serum/LIF (56.2±8.0% of Nanog:VNP expressing cells; Abranches et al., 2013).

4.2 Analysis of Gene Expression in Pluripotency

Pluripotency is maintained by a core network of TFs composed of Oct4, Sox2 and Nanog. In mESCs cultured in Serum/LIF, Oct4 and Sox2 have been described to be expressed homogeneously, while Nanog expression as being very heterogeneous. This heterogeneity allows the definition of two populations – Low- and High-Nanog, which show functional and molecular differences (Abranches et al., 2014).

The first experiments done for this MSc thesis concerned the re-analysis of *Nanog* and *Sox2* expression in mESCs cultivated in Serum/LIF, through smRNA-FISH. This method gives single cell resolution, and the quantification of mRNAs in a great number of cells allows the measurement of population dynamics.

4.2.1 *Nanog* and *Sox2* mRNA Expressions in mESCs

To re-assess the previously observed heterogeneous *Nanog* expression and homogeneous *Sox2* expression (Chambers et al., 2007; Singh et al., 2007; Abranches et al., 2013), quantification of *Nanog* and *Sox2* mRNAs was performed. Data is presented as histograms in figure 4.1. While *Nanog* mRNAs show a long-tailed distribution, with a high frequency of cells with low expression and few cells expressing high levels of transcripts, *Sox2* mRNAs reveal a Gaussian-like distribution, with many cells expressing an average number of transcripts and fewer cells expressing higher or lower transcript counts (Figure 4.1).

Examining the dispersion plot (Figure 4.1), it is perceptible a correlation between the two genes, confirmed by a spearman correlation coefficient of 0.69, as cells with high *Nanog* expression also have high *Sox2* expression.

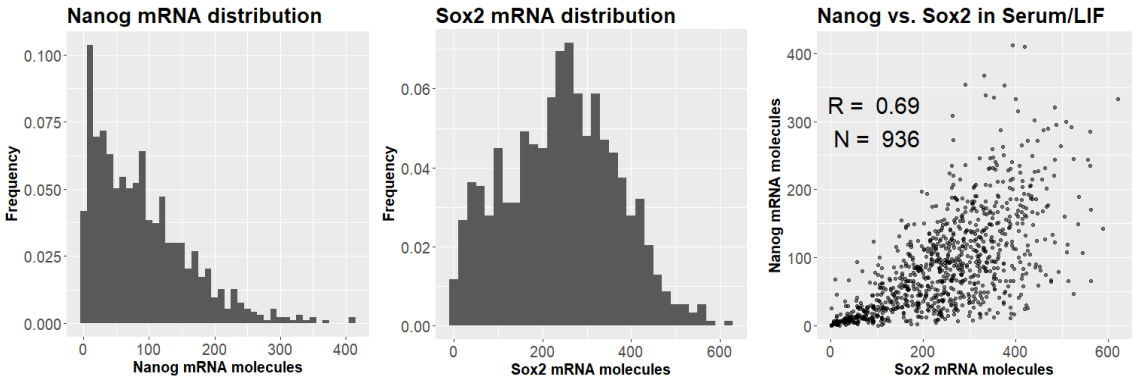

Figure 4.1 – *Nanog* and *Sox2* Transcript Distributions and their Correlation. Histograms of *Nanog* and *Sox2* distribution and dispersion plot with their correlation. Histograms bin width of 10 and 20 transcripts for *Nanog* and *Sox2*, respectively. R: Spearman correlation coefficient, N: number of cells analysed. In the dispersion plot each dot represents one mESC. smRNA-FISH performed in E14 mESCs.

Both genes exhibit a wide range of transcripts per cell, varying from 0 to 412 transcripts per cell in the case of *Nanog*, and from 1 to 621 transcripts for *Sox2*. The average number of transcripts per cell is 89 (± 72) for *Nanog*, and 243 (± 125) for *Sox2* (Table 4.1). *Nanog* median is lower than the mean, indicating a higher frequency of cells with fewer transcripts than the mean.

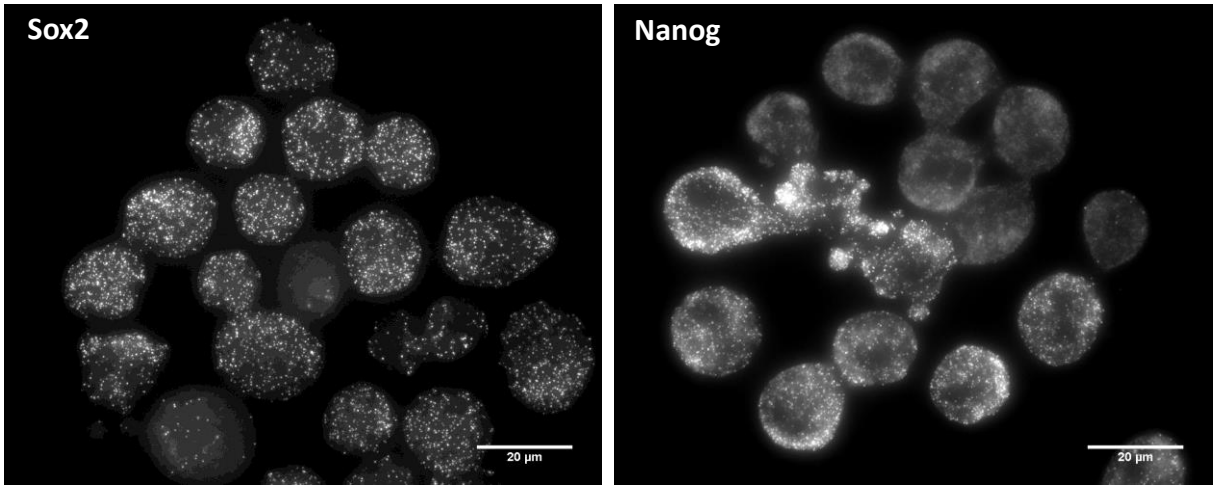

Figure 4.2 – *Sox2* and *Nanog* Expressions in Serum/LIF. Nanog heterogeneous expression compared to Sox2 more homogeneous expression in mESCs cultivated in Serum/LIF. Scale bars indicates 20µm.

FFs of both genes are much higher than 1 (58.59 and 64.30 for *Nanog* and *Sox2*, respectively), indicating that transcription occurs in bursts for both genes. CV values are 0.81 and 0.51, for *Nanog* and *Sox2*, respectively, implying high cell-to-cell variation in gene expression, with *Nanog* being more variable than *Sox2*.

Nanog and *Sox2* have mRNA half-lives of 3.7±0.9h and 1.8±1.0h, respectively (Abranches et al., 2013). As *Sox2* mRNAs reveal a Gaussian-like mRNA distribution but high FF and CV, this short mRNA half-life implies that *Sox2* transcriptional bursting must have a high frequency to generate the observed mRNA distribution. On the other hand, *Nanog* mRNA has a longer half-life, a high FF and a long-tailed distribution, indicating that *Nanog* transcriptional bursts are more spaced in time. This is in concordance with a previous study that showed that Nanog is mainly in an inactive transcription (OFF) state, with short transcription bursts (Ochiai et al., 2014).

Table 4.1 – Statistical Measurements for the Pluripotency Genes. For both genes, mean, median, standard deviation (SD), variance, Fano factor (FF), coefficient of variation (CV), minimum (Min) and maximum (Max) mRNA counts and number of ESCs analysed (N) are presented.

Gene	Mean	Median	SD	Variance	FF	CV	Min	Max	N
Nanog	88.62	75	72.05	5191.91	58.59	0.81	0	412	936
Sox2	242.89	247	124.97	15618.70	64.30	0.51	1	621	

The distinction between cells expressing high and low levels of transcripts relies on the definition of a threshold for each of the analysed genes. Previous work in the laboratory analysing *Nanog* and *Sox2* proteins and mRNAs expression in different culture conditions led to the definition of a threshold of 50 mRNAs to distinguish between Low- and High-Nanog or Sox2 states.

Using this threshold to distinguish the low and high-expressing populations, 62.71% of the cells were classified as having high *Nanog* expression and 37.27% as having low *Nanog* expression, while 92.52% of cells were classified has having high *Sox2* expression (Table 4.2). The correlation between expression of both genes reveals that 62.50% of the cells are in a state of pluripotency, exhibiting high expression of both genes. Only a very small portion of *Nanog*-expressing cells (0.21%) shows reduced *Sox2* expression, while 7.26% of mESCs do not express any of the genes. These cells are likely to be

differentiating, a phenomenon known to occur in Serum/LIF cultures. On the contrary, the 30% of cells which have high expression of *Sox2* and low expression of *Nanog* are hypothesised to be in a primed state of pluripotency. This is also suggested by the fact that 80.52% of Low-Nanog cells also exhibit expression of *Sox2,* implying that these are still within the pluripotency window.

Table 4.2 – Analysis of *Nanog* and *Sox2* Expressions. Number of cells in each subpopulation (N), percentage of cells relatively to the total number of cells (% to Total), to Low- or High-Nanog state (% to Nanog), and global percentages of Low- and High-Nanog (Nanog %) and Sox2 (Sox2 %) are presented. Cell were classified using the defined threshold of 50 transcripts for both genes.

Nanog	*Sox2*	N	% to Total	% to Nanog	Nanog %	Sox2 %
+	+	585	62,50	99,66	62.71	92.52
+	-	2	0,21	0,34		7.48
-	+	281	30,02	80,52	37.29	-
-	-	68	7,26	19,48		-

4.2.2 Priming Genes mRNA Expression on mESCs

In the primed state of pluripotency, the low/absent *Nanog* expression seems to correlate with concomitant expression of genes involved in lineage fate decisions, reflecting a state of lineage priming during which mESCs can exit the pluripotent state and initiate lineage differentiation (Martinez-Arias and Brickman, 2011; Abranches et al., 2014).

Previous work in the laboratory led to the identification, by RNA-Seq, of genes upregulated and downregulated in Low- and High-Nanog states (unpublished data). For this identification, mESCs were cultured in Serum/LIF and sorted based on Nanog:VNP fluorescence.

The genes with higher expression in Low-Nanog cells and residual or no expression in High-Nanog were termed “priming genes”. From the list of priming genes upregulated in Low-Nanog cells defined in the previous work, *Car2*, *Crabp2*, *Dnmt3b*, *Fgf5* and *Sox3* were selected for analysis as part of the MSc work here described.

Car2 (carbonic anhydrase 2) encodes a cytosolic zinc metalloenzyme, catalysing the reversible hydration of carbon dioxide (CO_2), being expressed in the embryonic placenta (Singh et al., 2005). *Crabp2* (cellular retinoic acid binding protein 2) encodes a cellular retinoic acid binding protein responsible for the transport of retinoic acid to its receptors in the nucleus. In embryos, it is expressed during neural development and is also increased during neural differentiation *in vitro* (Ruberte et al., 1992, Abranches et al., 2009). *Dnmt3b* encodes a *de novo* DNA methyltransferase essential for DNA methylation throughout mouse development (Okano et al., 1999). *Fgf5* (fibroblast growth factor 5) is expressed in the embryo in the post-implantation epiblast and is later restricted to the cells forming the three germ layers during gastrulation, being expressed only until their complete formation (Hébert and Martin, 1991). *Sox3* (sex determining region Y-box 3) encodes for a transcription factor involved in the activation of genes expressed in neural precursor cells, during neural differentiation (Abdelalim et al., 2014).

Expression of these genes was analysed by smRNA-FISH in mESCs grown is Serum/LIF. All genes exhibit long tail distributions, with low expression in most cells and few cells exhibiting high expression (Figure 4.3). The FFs of these genes vary from 5 to 90, indicating that their transcription occurs in bursts, although with different bursting dynamics. Furthermore, the high CVs of *Car2*, *Fgf5* and *Sox3* demonstrate that expression of these genes is highly variable. *Crabp2* and *Dnmt3b* are less variable than their counterparts, as revealed by their lower CVs.

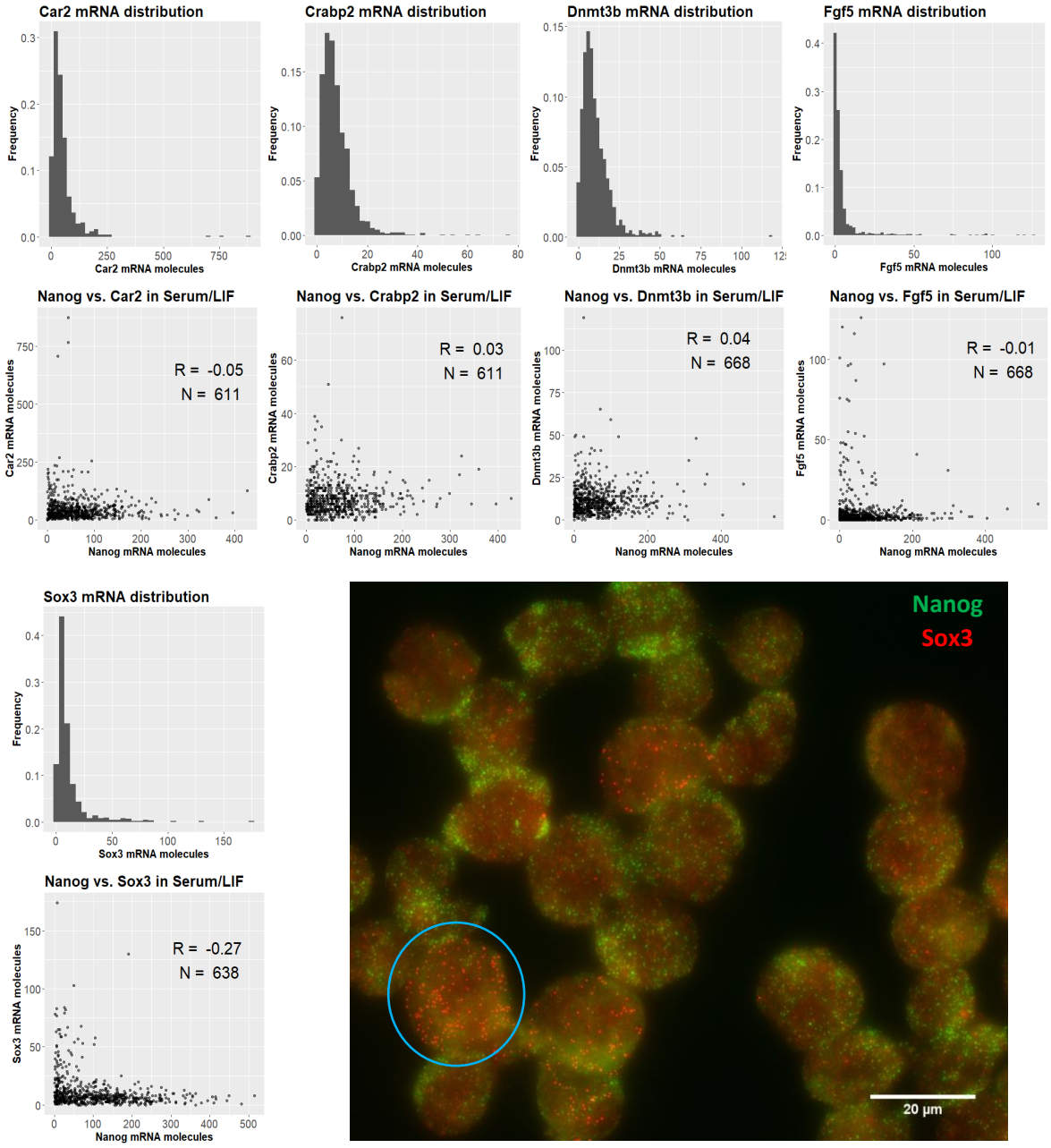

Figure 4.3 – Priming Genes Transcript Distributions and their Correlation to *Nanog*. Histograms of *Car2*, *Crabp2*, *Dnmt3b*, *Fgf5* and *Sox3*, and the dispersion plots with their correlation to *Nanog*. Histograms bin width of 20 for *Car2*, 2 for *Crabp2*, 2 for *Dnmt3b*, 5 for *Fgf5* and 5 for *Sox3*. R – Spearman correlation coefficient, N – number of cells analysed. In the dispersion plot each dot represents one mESC. Example of a smRNA-FISH experiment, imaging mRNAs of *Nanog* (green) and *Sox3* (red); Circle surrounding a primed mESC.

Examining the dispersion plots of priming gene correlation to *Nanog* (Figure 4.3), *Sox3* is the only gene exhibiting a negative correlation. Correlation coefficients of the other genes are 0, meaning that there is no correlation between their expression and Nanog expression. This might be due to the low number of priming cells (cells with low *Nanog* expression and high priming gene expression), which affects the calculation of the correlation parameters, and the high frequency of cells with low expression of both genes.

Although all of these candidate priming genes exhibit long-tailed distributions, the ranges of transcripts per cell vary, with *Car2* showing the highest expression level, up to a maximum of 875 transcripts per cell. The mean expression levels are also different among the candidate priming genes, with *Car2* having a mean of 50 (±67) transcripts per cell, *Fgf5* having only 5 (±13) transcripts per cell, and the remaining genes expressing 8 to 11 transcripts per cell (Table 4.3).

Table 4.3 – Statistical Measurements for the Analysed Priming Genes. For all genes, mean, median, standard deviation (SD), variance, Fano factor (FF), coefficient of variation (CV), minimum (Min) and maximum (Max) mRNA counts and number of ESCs analysed (N) are presented.

Gene	Mean	Median	SD	Variance	FF	CV	Min	Max	N
Car2	50,02	36	66,94	4481,42	89,59	1,34	0	875	611
Crabp2	8,04	7	6,43	41,40	5,15	0,80	0	76	1524
Dnmt3b	10,89	9	8,86	78,52	7,21	0,81	0	119	1085
Fgf5	5,30	2	12,76	162,71	30,71	2,41	0	126	1085
Sox3	11,26	7	15,71	246,78	21,91	1,39	0	174	638

4.3 Expression Patterns in Low- and High-Nanog:VNP Sorted mESCs

The two Nanog populations are molecularly and functionally different, and their differential gene expressions has been characterized (Abranches et al., 2014). To confirm the previous findings of higher priming gene expression in Low-Nanog cells, obtained by RNA-seq (unpublished data), expression of *Nanog*, *Fgf5* and *Sox3* expression was analysed in purified populations of Low-Nanog:VNP and High-Nanog:VNP cells, obtained by FACS sorting of Nd cells, cultured in Serum/LIF. Nd mESCs express a fluorescent protein (VNP) under regulation of the Nanog promoter, with similar half-lives to Nanog mRNA and protein, mimicking *Nanog* expression.

Analysing the histograms (Figure 4.4), it is visible a clear separation of the two Nanog populations. As expected, Low-Nanog:VNP cells exhibit a high frequency of cells with low *Nanog* transcript counts and High-Nanog:VNP cells exhibit higher frequency of cells with high *Nanog* transcript counts. This results in mean transcript counts of 26 (±32) and 108 (±69) mRNAs per cell in Low-Nanog:VNP and High-Nanog:VNP cells, respectively.

As for *Fgf5* and *Sox3* expression, Low-Nanog:VNP cells exhibit a higher expression of these genes (7±17 and 16±18 mean transcripts per cell of *Fgf5* and *Sox3*, respectively), with residual expression in High-Nanog:VNP cells (2±6 and 4±4 mean transcripts per cell of *Fgf5* and *Sox3*, respectively).

These results confirm at the single-cell level the global RNA-seq data (performed in populations), in which the selected priming genes were found to be more expressed in Low-Nanog:VNP cells.

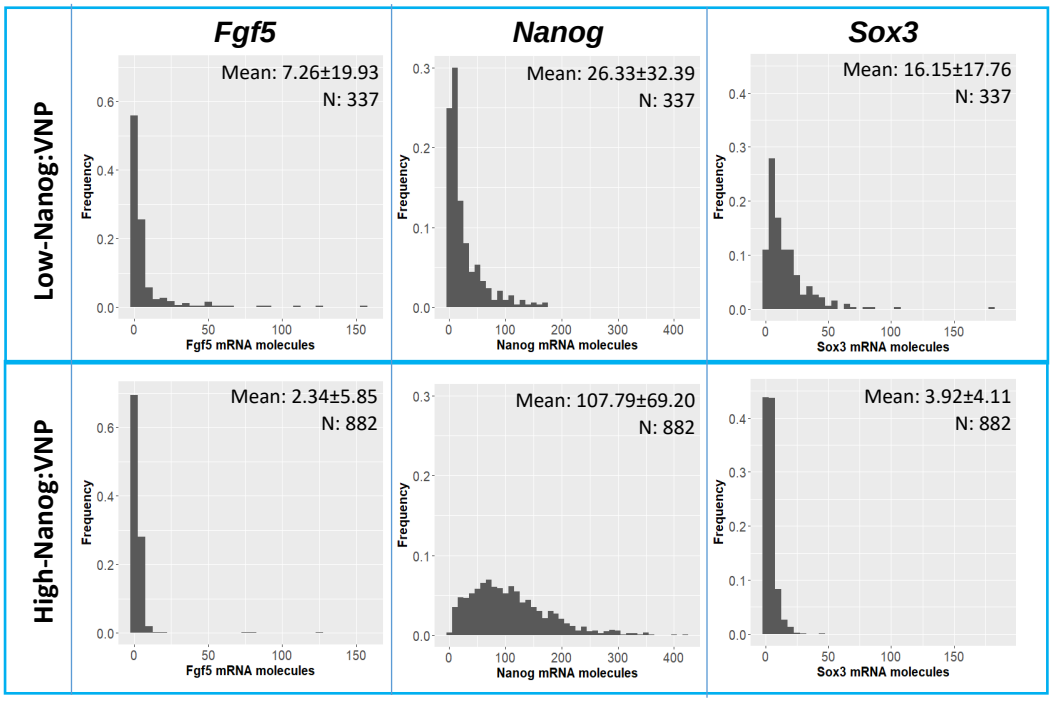

Figure 4.4 – *Fgf5*, *Nanog* and *Sox3* Expression in Low- and High-Nanog:VNP mESCs. Equal axis scales for better comparison; Bin width of 5 for *Fgf5* and *Sox3* and 10 for *Nanog*. For each experiment, the mean number of transcripts and number of cells analysed (N) is depicted in the histograms.

Applying the previously defined threshold for *Nanog* expression (Table 4.4; high expression above 50 mRNAs per cell), 16.62% of Low-Nanog:VNP cells can be classified as having high expression of *Nanog* mRNAs, while 20.52% of High-Nanog:VNP cells express less than 50 *Nanog* mRNAs. These discrepancies between expression of Nanog:VNP reporter and *Nanog* mRNAs can be due to the different time delays between transcription and mRNA translation/protein accumulation:

– Low-Nanog:VNP cells with >50 *Nanog* mRNA transcripts are likely to be cells in which *Nanog* transcription has started recently but in which the VNP reporter protein had not yet time to fold properly and be detectable (around 40');

– High-Nanog:VNP cells with <50 *Nanog* mRNA transcripts would be cells that still have VNP protein but have switched-off Nanog:VNP transcription. Although Nanog:VNP protein decay (1/2 life of ~2h) occurs faster than Nanog mRNA decay (1/2 life of ~4h), the number of protein molecules is much higher than mRNA molecules (>1.000 at least, see Schwanhausser et al, 2011), and one can postulate that cells that have switched-off *Nanog* transcription will keep detectable protein (Nanog or fluorescent VNP) for longer than mRNA.

Table 4.4 – Number and Percentage of mESCs Analysed in Low- or High-Nanog:VNP ESCs. Low- and High-Nanog defined by mRNA with a threshold of 50 mRNA transcrips; N: number of cells analysed; %: percentage of cell relatively to total.

	Low-Nanog:VNP		**High-Nanog:VNP**	
	Low-Nanog	High-Nanog	Low-Nanog	High-Nanog
N	281	56	181	701
%	83.38	16.62	20.52	79.48

4.3.1 Defining High Priming Gene Expression

Analysis of *Fgf5* and *Sox3* expression in Low- and High-Nanog:VNP cells confirmed different expression levels between the two cell populations (Figure 4.4), with Low-Nanog:VNP cells expressing higher levels of *Fgf5* and *Sox3* transcripts than High-Nanog:VNP cells, which show residual expression of these genes. We took advantage of the values obtained for this residual expression to establish a threshold level for defining a "low priming gene expression" state in which active transcription of these genes is unlikely to occur. mESCs with expression levels of priming genes above this threshold will be considered as transcriptionally active and classified in our experiments as "high priming gene expression". Thus, cells were classified as high *Fgf5* expression when they have above 10 transcripts per cell, and high *Sox3* expression above 20 transcripts per cell. Using these thresholds, only 1.24% and 0.91% of High-Nanog:VNP cells are classified as high expression of *Fgf5* and *Sox3*, respectively (dispersion plots on Supplementary Figure 7.1).

4.4 Priming Gene Expression in Pluripotency

The expression thresholds defined for *Fgf5* and *Sox3* (section 4.3.1) were used to classify high and low-expressing cells from cultured mESCs in normal Serum/LIF conditions (mESCs from 4.2.2). The results show that approximately 10% of these cells exhibit high expression of the two priming genes. We therefore used a similar value to define expression thresholds for the other selected priming genes *Car2*, *Crabp2* and *Dnmt3b*, taking into consideration the distribution values obtained for each gene in normal Serum/LIF mESC cultures. The chosen threshold values were 100, 20 and 15 mRNAs per cell for respectively *Car2*, *Crabp2* and *Dnmt3b* (Supplementary Table 7.1).

Using these threshold values to analyse the percentage of cells with high expression of each priming gene relatively to *Nanog* mRNA level (Table 4.5), the results show a higher expression of each priming gene in the Low-Nanog state. The larger difference is found for *Sox3*, as 25.21% of the Low-Nanog cells are expressing high levels of *Sox3*, whereas only 5.45% of High-Nanog cells present high *Sox3* expression. Similar results are also found for the other priming genes, although with smaller differences between *Nanog* states.

Table 4.5 – Percentage of mESCs with High Expression of Priming Genes Relatively to *Nanog* Level. Low- and High-Nanog defined by mRNA expression.

% High-Car2		**% High-Crabp2**		**% High-Dnmt3b**		**% High-Fgf5**		**% High-Sox3**	
Low-Nanog	High-Nanog	Low-Nanog	High-Nanog	Low-Nanog	High-Nanog	Low-Nanog	High-Nanog	Low-Nanog	High-Nanog
13.88	6.06	12.10	8.48	14.41	10.37	13.53	6.40	25.21	5.45

4.5 Nanog, Tet1 and PRC2 in Priming Gene Regulation

Previous work in the laboratory led to the identification of genes upregulated or downregulated in Low- and High-Nanog states (unpublished data). For this identification, mESCs were cultured in Serum/LIF and sorted based on Nanog:VNP fluorescence. A list of genes with significantly higher expression in Low-Nanog:VNP cells or with higher expression on High-Nanog:VNP cells was analysed through Enrichr (Chen et al., 2013; Kuleshov et al., 2016), a chromatin enrichment analysis tool based on Chip-Seq results, to identify proteins associated with the genes upregulated in each of the *Nanog* expression states. Results showed that genes upregulated in High-Nanog cells are bound by proteins involved in the pluripotency network, whereas genes upregulated in the Low-Nanog state are mainly bound by members of the PRC2 complex (unpublished data).

Nanog has been shown interact with Tet1 (Costa et al., 2013) and Ezh2, a PRC2 component (Gagliardi et al., 2013). Tet1 is the enzyme responsible for demethylation of 5mC into 5hmC, and Ezh2 is the catalytic subunit with methyltransferase activity of the PRC2 complex, responsible for the methylation of H3K27. Furthermore, it is known that PRC2 binds preferentially to 5hmC enriched CpG regions (CpG islands) in the promoters of developmentally relevant genes (Wu et al., 2011), indicating that there is a connection between Nanog, Tet1 and PRC2.

We therefore hypothesised that Nanog acts to repress priming gene expression in pluripotent mESCs through recruitment of Tet1, which will create hypomethylated CpG regions nearby Nanog targets. These hypomethylated CpG regions would be subsequently recognized by the PRC2 complex, whose activity will lead to increased methylation of H3K27 and consequent repression of Nanog target genes.

To test this hypothesis, mESCs were exposed to chemical compounds that interfere with the activity of Tet1 and PRC2, namely ascorbic acid (AA) and GSK343, for 48h (Figure 4.5). AA is known to stimulate Tet1 activity, while GSK343 is an inhibitor of Ezh2 (Verma et al., 2012; Blaschke et al., 2013; Yin et al, 2013; Hore et el., 2016). Our prediction to be tested is that stimulation of Tet1 activity by AA should increase 5hmC levels near Nanog target priming genes, and therefore their repression via PRC2; on the contrary, inhibition of Ezh2/PRC2 should block deposition of H3K27me3 in Nanog targets (priming genes) and result in their increased expression (due to lack of repressive H3K27me3), in a Nanog-dependent manner. Addition of both AA and GSK343 should result in increased 5hmC levels near priming genes, but inhibition of Ezh2 should prevent the subsequent H3K27me3 associated repression, reverting priming gene expression to control levels.

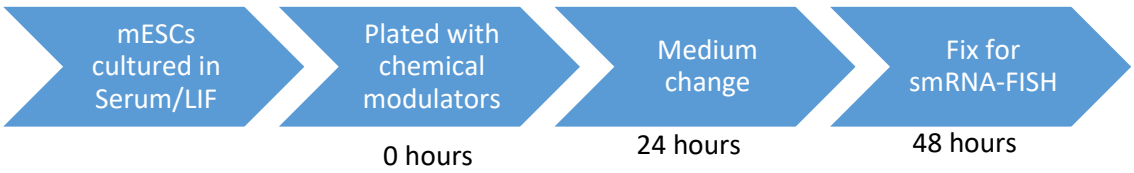

Figure 4.5 – Experimental Design of Chemical Modulators Assay. mESCs were cultured in Serum/LIF for three passages before the addition of the modulators (time 0h), and further cultured in the presence of the modulators for 48h, with a medium change at 24h. At 48h, Nanog:VNP levels were assessed, mESCs were fixed and smRNA-FISH was performed.

4.5.1 Morphology and Nanog:VNP Dynamics

Morphology of mESCs and Nanog:VNP levels were assessed at the end of mESC culture (48h). mESC cultured in the presence of DMSO had similar morphology and Nanog:VNP levels to mESCs cultured in Serum/LIF (section 4.1). The addition of GSK343 caused no alteration in morphology, and a small increase in Nanog:VNP expression (an average of 67% of Nanog:VNP expressing cells versus 61% in DMSO, three replicates). Unlike GSK343, AA treatment, either alone or in combination with GSK343, caused morphological changes in mESC cultures, decreasing the number of peripheral cells with irregular forms and increasing the clusters' roundedness. This more pluripotent morphology is accompanied with an increase in Nanog:VNP levels (from 61% of Nanog:VNP expressing cells in DMSO to 79% in AA, and 82% in AA+GSK343; Table 4.7).

4.5.2 Analysis of Gene Expression by smRNA-FISH

Expression of the following genes were analysed by smRNA-FISH in the different experimental conditions: *Nanog*, *Crabp2*, *Fgf5* and *Sox3*. The results are depicted as histograms in Figure 4.6 and statistical measurements in Table 4.6.

Analysis of these histograms reveal that, in the control condition (DMSO), there are no alterations in the mRNA distribution relatively to mESCs cultured only in Serum/LIF (without DMSO – section 4.2.2). This is an indication that DMSO treatment does not affect the expression of the analysed genes, with the exception of Car2, that presents higher expression levels (mean of 121 transcripts per cell in DMSO versus 50 transcripts per cell in Serum/LIF).

The expression of *Crabp2*, *Fgf5* and *Sox3* at the population level is not significantly affected by any of the treatments (GSK343, AA or AA+GSK343). *Nanog* expression, on the other hand, is affected by exposure to AA (mean of 115 and 112 transcripts per cell in AA and AA+GSK343 versus 86 in DMSO), confirming that the higher protein expression, measured as Nanog:VNP levels, is accompanied by higher *Nanog* mRNA expression. This higher expression of *Nanog* is revealed in the distribution histograms by the decreased frequency of cells with low transcript counts and an increase of cells with high transcript counts. This result indicates that Tet1 is regulating Nanog expression, as previously shown (Ito et al., 2010).

Car2 also seems to be Tet1-regulated, increasing its expression in cells exposed to AA (mean of 150 transcripts per cell versus 121 in DMSO). This effect is diluted when cells are also exposed to GSK343 (mean of 129 transcripts per cell versus 150 in AA) suggesting that besides Tet1, Ezh2/PRC2 might also be involved in *Car2* regulation.

The lack of apparent regulation of the other genes at the population level does not mean absence of regulation by Tet1 and PRC2, as it is expectable that effects of the modulators occur only in a small number of cells, those that have started to express *Nanog* during the 48h culture period (transiting from the Low-Nanog state). In cells that are stably High-Nanog during this 48h period, no changes in H3K27me3 on Nanog targets are expected, as the GSK343 inhibitor will not promote H3K27 demethylation, preventing only de novo H3K27 methylation when cells transit from Low- to High-Nanog state. Only in these cells, the inhibition should result in lack of Nanog/Tet1/Ezh2 activity and lead to absence of priming gene repression. To measure and quantify these effects, analysis at single cell level is required, correlating priming gene expression with the *Nanog* expression state.

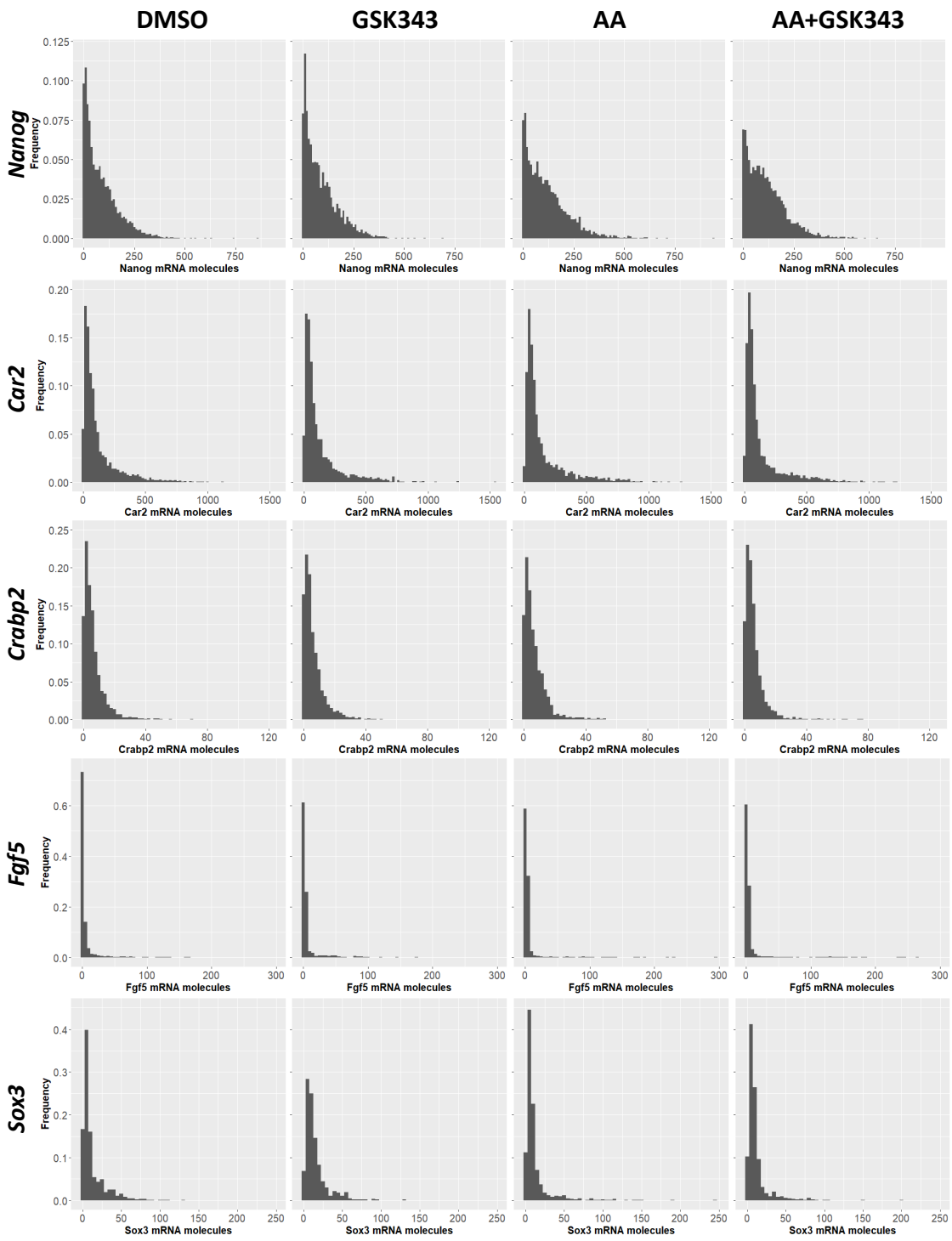

Figure 4.6 – Histograms Representing the Expression of *Nanog*, *Car2*, *Crabp2*, *Fgf5* and *Sox3* in mESCs treated with the Chemical Modulators. Distribution of gene expression in ESCs treated with DMSO, GSK343, AA and AA+GSK343; Bin width of 10 for *Nanog*, 20 for *Car2*, 2 for *Crabp2* and 5 for *Fgf5* and *Sox3*.

Table 4.6 – Statistical Measurements for the Analysed Genes in mESCs treated with the chemical modulators. For all genes, mean, median, standard deviation (SD), variance, Fano factor (FF), coefficient of variation (CV), minimum (Min) and maximum (Max) mRNA counts and number of ESCs analysed (N) are presented.

	Gene	Mean	Median	SD	Variance	FF	CV	Min	Max	N
DMSO	*Nanog*	85.94	63	84.23	7095.43	82.56	0.98	0	865	4241
	Car2	121.03	69	145.00	21025.37	173.72	1.20	0	1122	3380
	Crabp2	6.99	5	7.86	61.73	8.84	1.12	0	121	
	Fgf5	5.63	1	16.83	283.21	50.27	2.99	0	164	861
	Sox3	12.59	6	16.20	262.58	20.86	1.29	0	131	
GSK343	*Nanog*	89.00	66	84.20	7089.50	79.66	0.95	0	695	2392
	Car2	129.77	67	163.14	26614.31	205.08	1.26	0	1546	1719
	Crabp2	6.64	5	6.60	43.50	6.55	0.99	0	50	
	Fgf5	7.10	2	18.33	336.12	47.35	2.58	0	173	673
	Sox3	14.66	10	14.70	216.07	14.74	1.00	0	130	
AA	*Nanog*	115.46	91	104.68	10957.26	94.90	0.91	0	942	2646
	Car2	149.71	78	176.92	31301.90	209.08	1.18	0	1253	2115
	Crabp2	7.33	5	7.26	52.66	7.19	0.99	0	53	
	Fgf5	7.13	2	24.48	599.10	84.02	3.43	0	294	850
	Sox3	12.55	7	20.63	425.66	33.91	1.64	0	243	
AA+GSK343	*Nanog*	111.84	94	93.65	8771.25	78.42	0.84	0	658	3649
	Car2	128.80	66	160.95	25905.75	201.13	1.25	0	1216	2719
	Crabp2	6.36	5	6.40	40.90	6.43	1.01	0	76	
	Fgf5	8.17	2	27.64	764.11	93.49	3.38	0	265	1012
	Sox3	11.35	7	14.94	223.34	19.68	1.32	0	200	

4.5.3 Analysis of Priming Gene Expression in Low- and High-Nanog mESCs

To assess the effect of the modulators in priming gene expression in Low- and High-Nanog populations, a single cells analysis was performed (Table 4.7).

In control conditions (DMSO) there is a higher number of cells with high expression of each of the priming genes in cells classified as Low-Nanog (less than 50 Nanog transcripts), than High-Nanog. These results are similar to those previously obtained in Serum/LIF cells (section 4.2.2).

Car2

Analysis of *Car2* expression reveals no alteration in cells treated with GSK343, and an increase in the percentage of High-Nanog cells with high expression of *Car2* when exposed to AA (32.94% of cells in AA versus 23.04% in DMSO). However, this increase is also verified in Low-Nanog cells (63.17% of cells in AA versus 51.53% in DMSO), suggesting that *Car2* regulation by Tet1 should be Nanog-independent. Although GSK343 alone had no effect on expression of this gene, neither in Low- nor in High-Nanog cells, it prevents the AAs' increase in *Car2* expression. This suggests that *Car2* expression in mESCs is regulated by the combined activity of Tet1 and PRC2, but this is independent of Nanog expression.

Crabp2

Similar results are found for *Crabp2*, with GSK343 having no effect on the expression of this gene. AA also seems to cause a small increase in expression of *Crabp2* in both Low- and High-Nanog cells, reverted when GSK343 is added with AA.

Fgf5

Unlike *Car2* and *Crabp2*, GSK343 treatment causes a slight increase of cells with high expression of *Fgf5*, although this increase seems to be Nanog-independent, as it is detected in both High-Nanog (8.82% of cells with high *Fgf5* expression versus 7.30% in DMSO) and Low-Nanog (15.94% of cells with high *Fgf5* expression versus 12.71% in DMSO) cells. An opposite effect is observed when mESCs are treated with AA, with a decrease of cells with high *Fgf5* expression in both *Nanog* states (from 7.30% and 12.71% in DMSO, to 6.22% and 9.23% in AA, in High-Nanog and Low-Nanog, respectively). The repressive effect of AA is reverted when cells are also cultured in the presence of GSK343, with a small increase of High-Nanog cells with high *Fgf5* expression (8.39% in AA+GSK343 versus 6.22% in AA and 7.30% in DMSO), to similar levels of cells treated with GSK343 (8.82%).

These results suggest that inhibition of H3K27me3 deposition by GSK343 causes an increase of cells with high *Fgf5* expression, while the increase of 5hmC by Tet1/AA causes a decrease in the frequency of these cells. Furthermore, analysis of cells treated with both modulators indicates that the repression of *Fgf5* expression by AA is dependent of Ezh2/PRC2 activity. These effects might be independent of Nanog, as the increase and the decrease caused by GSK343 and AA, respectively, are found in cells in both states. Strikingly, the effect of the combination of modulators caused only an increase of High-Nanog cells with high expression of *Fgf5*, not affecting Low-Nanog cells.

Sox3

Analysis of *Sox3* expression reveals that there is an increase of High-Nanog cells with high *Sox3* expression in GSK343 treated cells (13.35% in GSK343 versus 9.51% in DMSO), with no alteration in Low-Nanog cells (32.25% in GSK343 versus 33.01% in DMSO). In AA treated mESCs, the frequency of High-Nanog cells with high *Sox3* expression decreases (from 9.51% in DMSO to 6.91% in AA), and a small decrease in Low-Nanog cells with high *Sox3* expressing cells is also observed (33.01% in DMSO to 26.20% in AA). These results indicate that *Sox3* expression is regulated by PRC2 and Tet1, in a Nanog-dependent manner.

Unexpectedly, treatment of cells with both modulators did not revert the repressive effect of AA (6.29% of High-Nanog cells with high *Sox3* expression in AA+GSK343 versus 6.91% in AA). This result indicates that although PRC2 and Tet1 regulate *Sox3* in a Nanog-dependent manner, they do not seem to act together, as the proposed model hypothesises. This result contrasts with previous results from the laboratory that performed the same experiment. Those results showed that *Sox3* expression was restored

when cells were treated with both modulators, indicating that *Sox3* was regulated by Tet1 and PRC2, in a Nanog-dependent manner, with PRC2 acting downstream of Tet1. To clarify this unexpected result, *Sox3* regulation is further analysed in the next section (section 4.6).

Table 4.7 – Percentage of mESCs with High Expression of Priming Genes Relatively to Nanog Level. Low- (LN) and High-Nanog (HN) defined by mRNA expression, and the percentage of High-Nanog cells measured by mRNA and Nanog:VNP.

	% H-Car2		% H-Crabp2		% H-Fgf5		%H-Sox3		% High-Nanog (mRNA)	% High-Nanog (VNP)
	LN	HN	LN	HN	LN	HN	LN	HN		
DMSO	51.53	23.04	16.88	5.46	12.71	7.30	33.01	9.51	55.58	60.95
GSK343	52.13	25.60	16.64	6.05	15.94	8.82	32.25	13.35	58.07	67.00
AA	63.17	32.94	22.71	8.30	9.23	6.22	26.20	6.91	73.19	79.13
AA + GSK343	59.44	25.18	18.44	4.16	11.65	8.39	24.50	6.29	75.15	82.47

4.6 Tet1 and PRC2 in Nanog-state Transitions

Here, it is hypothesised that Nanog regulates the repression of priming genes by recruiting Tet1 to their regulatory regions and converting 5mC into 5hmC. This hypomethylated region will be recognized by PRC2, that will trimethylate H3K27 and repress the expression of priming genes.

As the modulators, AA and GSK343, are expected to act in a Nanog-dependent manner, increasing *de novo* demethylation of 5mC and methylation of H3K27, respectively, mESCs were sorted into Low- and High-Nanog populations and cultured in the presence of the inhibitors (AA, GSK343, or both) for 48h. Our prediction is that the effect of the modulators should be more evident in mESCs transitioning from the Low-Nanog state into High-Nanog state, in which Nanog is present and priming genes must be repressed.

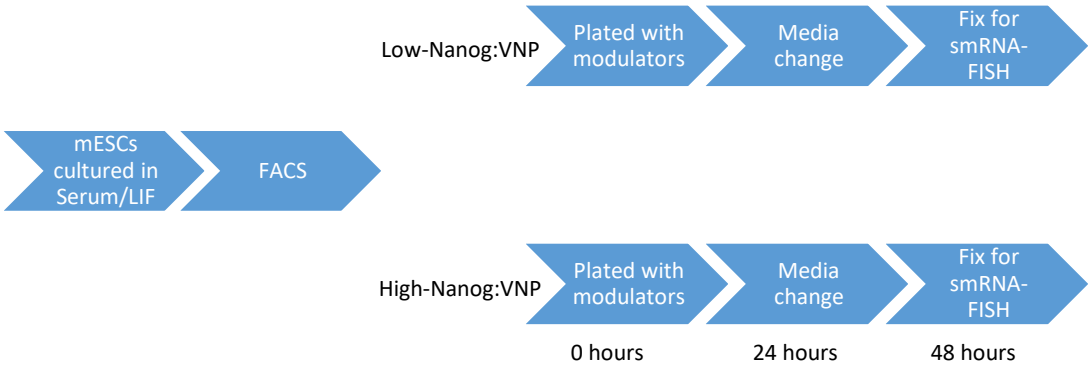

Figure 4.7 – Experimental Design of Sorted mESCs treated with the Chemical modulators. mESCs were expanded in Serum/LIF until an appropriate number of cells. Low- and High-Nanog populations were sorted based on Nanog:VNP fluorescence and plated in Serum/LIF supplemented with the chemical modulators: GSK343, AA, both and DMSO as a control. 24 hours post-plating, medium was change to assure the concentration of the modulators and, at 48h, Nanog:VNP levels were assessed and mESCs fixed for smRNA-FISH. FACS – Fluorescence-Activated Cell Sorting.

4.6.1 mESCs Morphology and Nanog:VNP Dynamics

The sorted Low- and High-Nanog:VNP cells were plated in Serum/LIF medium supplemented with the chemical modulators for 48h. At the end of this period, morphology was assessed, Nanog:VNP levels were measured by flow cytometry and mESCs were fixed for smRNA-FISH (Figure 4.7, Supplementary Figure 7.2).

The two cell populations (Low-Nanog:VNP and High-Nanog:VNP) exhibit different morphologies, as the High-Nanog:VNP mESCs grow mainly in rounder clusters while Low-Nanog:VNP mESCs grow in flatter clusters, with many irregular cells in the periphery of the clusters. Low-Nanog:VNP cells exposed to AA (alone or combined with GSK343) also presented round clusters with less peripheric cells, similarly to High-Nanog:VNP control mESCs.

In control cells (DMSO, Table 4.10), Nanog:VNP is being expressed in 15.4% of Low-Nanog:VNP cells and 55.6% of High-Nanog:VNP cells. These results are in concordance with a previous study, which showed that although Low-Nanog:VNP cells are able to restore the normal heterogeneous expression, the necessary time is higher in Low-Nanog:VNP cells. While High-Nanog:VNP restore the initial heterogeneity in 2 days, Low-Nanog:VNP cells may take up to 4 days to do the same (Abranches et al., 2013). A possible explanation is that it takes longer to reach detectable high levels of Nanog:VNP when cells start to produce it during transition from the Low-Nanog:VNP state, when compared with the Nanog:VNP decay rate when cells transit in the opposite direction.

Treatment of Low-Nanog:VNP and High-Nanog:VNP mESCs with GSK343 has different effects based on the initial Nanog expression level (Table 4.10). Starting with Low-Nanog:VNP sorted cells cultured for 48h, the percentages of Low- and High-Nanog:VNP cells is identical after 48h in DMSO or GSK343 conditions (15.4% of cells with high Nanog:VNP expression versus 14.6% in DMSO). In the case of sorted High-Nanog:VNP mESCs, presence of GSK343 for 48 causes a slower emergence of Low-Nanog:VNP cells with more cells maintained in the High-Nanog:VNP state (67.4% of cells with high Nanog:VNP expression in GSK343 versus 55.6% in DMSO). This result is compatible with a previous study in iPSCs, in which it was shown that Ezh2 null mESCs exhibit higher expression of Nanog due to the expansion of the High-Nanog population (Villasante et al., 2011). GSK343 is, thus, a good inhibitor, mimicking the complete removal of Ezh2. As the effect is only found in High-Nanog:VNP cells, Ezh2/PRC2 should be part of the mechanisms contributing to Nanog repression. In the bulk population experiment (section 4.5), GSK343 effect was not as noticeable due to the presence of cells in which Nanog was not expressed (Low-Nanog cells).

In AA treated cells (Table 4.10), we observed that mESCs transit faster from the Low- to the High-Nanog:VNP state than in control DMSO treatment (49.9% of cells with high Nanog:VNP expression versus 14.6% in DMSO). Similar results were obtained when starting with High-Nanog:VNP cells, which do not transit as fast to the Low-Nanog state as DMSO treated mESCs (73.8% of cells with high Nanog:VNP expression versus 55.6% in DMSO). These results confirm our previous results (section 4.5) indicating Nanog self-regulation through Tet1, likely by maintaining its promoter in a hypomethylated and more active state.

When Low-Nanog:VNP and High-Nanog:VNP mESCs are cultured in the presence of both inhibitors, the results are similar to the effect of AA alone, with a cumulative effect of GSK343, resulting in an increase of Nanog:VNP in Low-Nanog:VNP cells, and a maintenance of higher expression levels of Nanog:VNP in High-Nanog:VNP cells (52.7% and 81.7% of cells with high Nanog:VNP expression in AA+GSK343 versus 49.9% and 73.8% in AA in Low- and High-Nanog:VNP, respectively; Table 4.10).

Our classification of cells based on the expression of *Nanog* mRNA allows the comparison between the number of cells classified as High-Nanog by mRNA and by Nanog:VNP expression. Although *Nanog* mRNA quantification provides, in most cases, similar results to Nanog:VNP, it seems that DMSO and

GSK343 have a striking effect in dissociating the two processes when cells transit from the Low- to the High-Nanog state. In both cases, the percentage of High-Nanog cells quantified through VNP is significantly lower than the percentage of High-Nanog cells classified by mRNA expression, after Low-Nanog:VNP cells were cultured for 48h. One possible explanation is that this effect is due to the presence of DMSO (GSK343 is diluted in DMSO), that might interfere with *Nanog* mRNA translation in mESCs that have activated Nanog transcription when transiting to a High-Nanog state.

4.6.2 Gene Expression in Low- and High-Nanog:VNP mESCs

Our previous analysis of Low-Nanog:VNP sorted cells showed that there is low *Nanog* expression (26 transcripts per cell) and high *Sox3* expression (mean of 16 transcripts per cell). The opposite is found in High-Nanog:VNP cells, which show high *Nanog* expression (mean of 108 transcripts per cell) and low *Sox3* expression (mean of 4 transcripts per cell; section 4.3). These mESCs correspond to mESCs in time 0h of this experiment.

Low-Nanog:VNP

Analysing the results of sorted Low-Nanog:VNP cells grown in the presence of the modulators for 48h (Figure 4.8 and Table 4.8), there is no global effect of GSK343 on *Sox3* expression, while *Nanog* expression increases (from a mean of 73 to 83 transcripts per cell) and *Car2* expression decreases (from a mean of 197 to 173 transcripts per cell).

AA treatment affects all the analysed genes, increasing *Nanog* (from a mean of 73 to 110 transcripts per cell) and *Car2* (from a mean of 197 to 225 transcripts per cell) expressions, while reducing *Sox3* expression (from a mean of 21 to 13 transcripts per cell), relatively to DMSO. This effect is also seen through visual analysis of the histograms, with an increased frequency of cells with higher *Nanog* expression and lower *Sox3* expression.

When Low-Nanog:VNP cells were treated with both modulators, part of AA effects are lost, with addition of GSK343 causing a partial rescue of AA effect on *Sox3*, increasing its expression (from a mean of 13 to 16 transcripts per cell, from AA to AA+GSK343).

These results are in concordance with the proposed model of priming gene regulation by Nanog, Tet1 and PRC2. The simultaneous presence of AA, increasing Tet1 activity, and GSK343 inhibiting Ezh2/PRC2, causes an increase of *Sox3* expression in Low-Nanog:VNP mESCs, indicating that PRC2 acts downstream of Tet1.

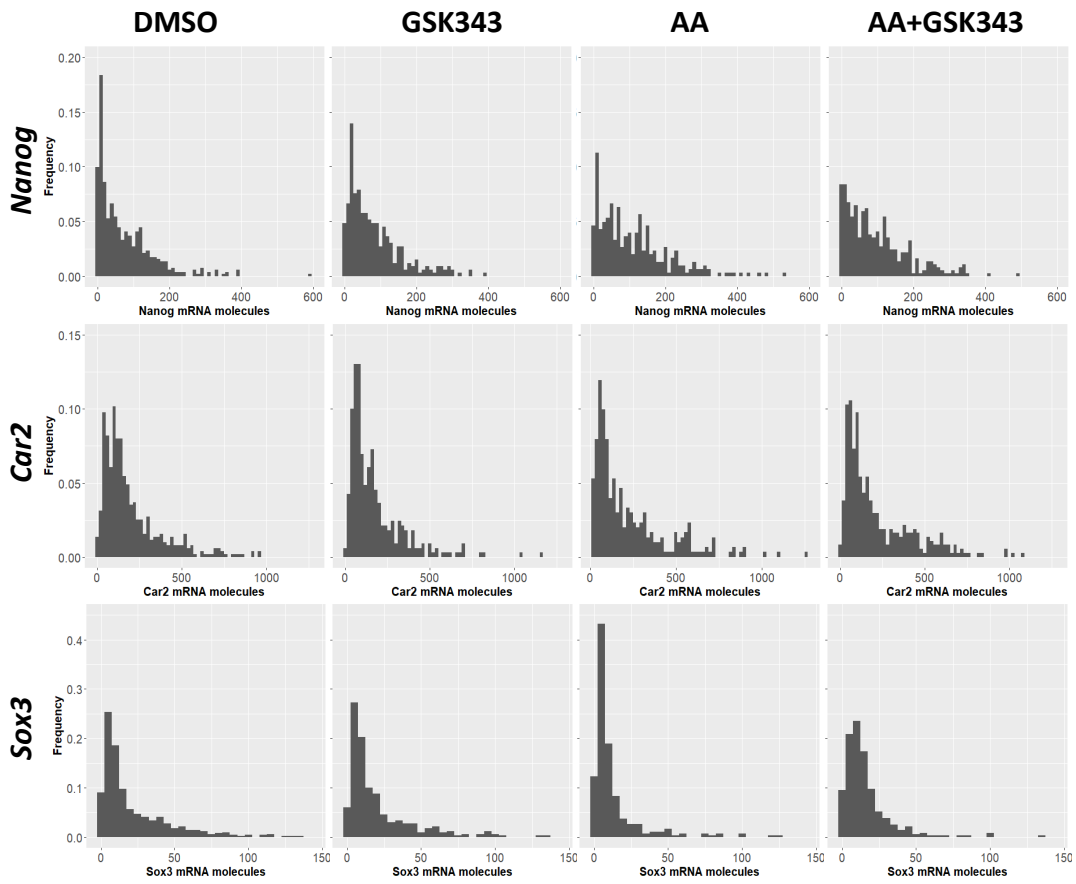

Figure 4.8 – *Nanog*, *Car2* and *Sox3* expression in Low-Nanog:VNP mESCs treated with the Chemical Modulators. Distribution of gene expression in mESCs treated with DMSO, GSK343, AA and AA+GSK343; Bin width of 10 for *Nanog*, 20 for *Car2* and 5 for *Sox3*.

Table 4.8 - Statistical Measurements for the Analysed Genes in Low-Nanog:VNP mESCs treated with the Chemical Modulators. For all genes, mean, median, standard deviation (SD), variance, Fano factor (FF), coefficient of variation (CV), minimum (Min) and maximum (Max) mRNA counts and number of ESCs analysed (N) are presented.

	Gene	Mean	Median	SD	Variance	FF	CV	Min	Max	N
DMSO	*Car2*	197.20	138.50	176.89	31291.49	158.68	0.90	0	961	512
	Nanog	73.16	46.00	78.47	6157.56	84.17	1.07	0	593	
	Sox3	21.36	11.00	23.97	574.64	26.91	1.12	0	133	
GSK343	*Car2*	173.42	121.00	165.03	27233.44	157.04	0.95	1	1153	330
	Nanog	83.14	61.50	75.14	5645.73	67.91	0.90	0	386	
	Sox3	20.45	12.00	22.91	524.89	25.66	1.12	0	137	
AA	*Car2*	225.16	141.00	217.29	47214.11	209.69	0.97	16	1254	301
	Nanog	110.04	86.00	97.00	9408.53	85.50	0.88	0	530	
	Sox3	12.83	7.00	18.27	333.78	26.02	1.42	0	123	
AA+ GSK343	*Car2*	220.33	140.00	209.91	44063.10	199.98	0.95	2	1275	369
	Nanog	96.95	73.00	86.68	7513.96	77.50	0.89	0	490	
	Sox3	16.08	11.00	17.36	301.46	18.75	1.08	0	143	

High-Nanog:VNP

Analysing the results of sorted High-Nanog:VNP mESCs cultured in the presence of the chemical modulators for 48h, no difference is found between DMSO and GSK343 relatively to *Car2* expression, while *Nanog* and *Sox3* become less expressed (decreasing from a mean of 119 and 15 to 107 and 10, respectively).

Culturing High-Nanog:VNP mESCs in the presence of AA led to an increase of *Nanog* transcription (from a mean of 119 mRNAs per cell in DMSO to 178 mRNAs per cell in AA) and a decrease of *Sox3* expression (from a mean of 15 transcripts per cell in DMSO to 7 transcripts per cell in AA).

When High-Nanog:VNP mESCs were cultured in the presence of AA+GSK343, we also observed a global increase in *Nanog* mRNA expression (from a mean of 119 mRNAs per cell in DMSO to 158 mRNAs per cell in AA+GSK343), while *Sox3* is not further affected (7 transcripts per cell in AA and 8 in AA+GSK343).

Car2 expression, on the other hand, is negatively affected by AA, when compared to DMSO (decrease from a mean of 106 to 91 transcripts per cell). This AA effect on *Car2* expression is lost when cells are also treated with GSK343 (increase to a mean of 104 transcripts per cell), indicating that *Car2* might also be regulated by Tet1 and PRC2, but only in cells with high *Nanog* expression. This effect was not seen in bulk cultures of non-sorted mESCs.

The great increase of Nanog expression that is also visible in the distribution (Figure 4.9) is an indication that Nanog self-regulation is, in part, done via Tet1. This effect is only seen in cells that started in High-Nanog:VNP and is not found in cells that started in Low-Nanog:VNP due to the required presence of Nanog protein (only present in High-Nanog:VNP cells).

Table 4.9 - Statistical Measurements for the Analysed Genes in High-Nanog:VNP mESCs treated with the Chemical Modulators. For all genes, mean, median, standard deviation (SD), variance, Fano factor (FF), coefficient of variation (CV), minimum (Min) and maximum (Max) mRNA counts and number of ESCs analysed (N) are presented.

	Gene	Mean	Median	SD	Variance	FF	CV	Min	Max	N
DMSO	***Car2***	105.79	89	77.22	5963.36	56.37	0.73	6	570	379
	Nanog	119.00	90	99.81	9962.25	83.71	0.84	0	557	
	Sox3	14.88	9	16.63	276.41	18.57	1.12	0	151	
GSK343	***Car2***	107.45	83	85.64	7334.11	68.26	0.80	1	677	364
	Nanog	107.25	93	84.71	7176.43	66.91	0.79	0	482	
	Sox3	9.56	5	12.06	145.51	15.22	1.26	0	81	
AA	***Car2***	90.64	73	71.66	5135.85	56.66	0.79	0	596	329
	Nanog	177.86	162	109.68	12030.08	67.64	0.62	2	620	
	Sox3	6.73	5	7.01	49.12	7.30	1.04	0	74	
AA+ GSK343	***Car2***	103.67	88	73.41	5389.74	51.99	0.71	5	625	251
	Nanog	158.05	132	105.28	11083.86	70.13	0.67	0	531	
	Sox3	8.24	6	8.55	73.02	8.86	1.04	0	66	

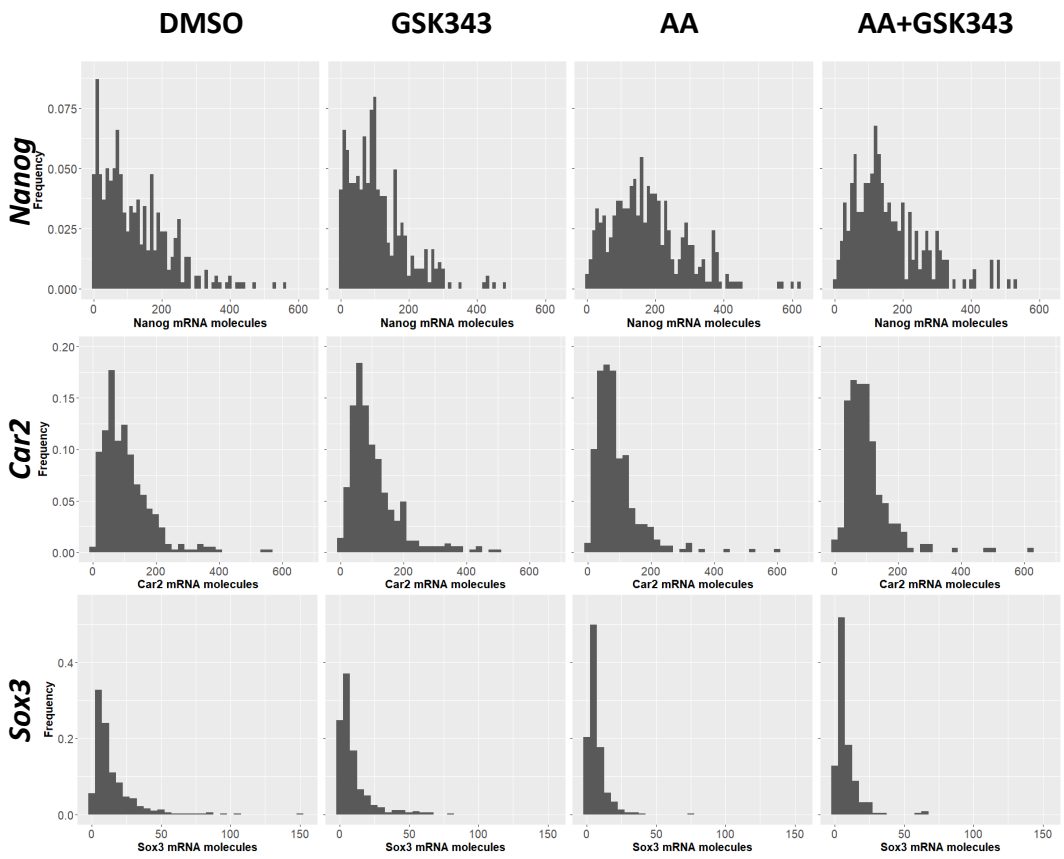

Figure 4.9 – *Nanog*, *Car2* and *Sox3* expression in High-Nanog:VNP mESCs treated with the Chemical Modulators. Distribution of gene expression in ESCs treated with DMSO, GSK343, AA and AA+GSK343; Bin width of 10 for *Nanog*, 20 for *Car2* and 5 for *Sox3*.

4.6.3 Analysis of Priming Gene Expression in Low- and High-Nanog ESCs

To assess the effect of the modulators in priming gene expression, single-cell analysis was performed, and the results correlated to Nanog mRNA expression; Table 4.10).

According to our initial hypothesis, it is expected that interference with the repressive activity of Nanog/Tet1/PRC2 on priming genes should become more evident when mESCs transit from the Low- to the High-Nanog state and control of priming genes is necessary. We thus focused our analysis on sorted Low-Nanog:VNP mESCs cultured for 48h with the different modulators, a time when ~50% of the cells have transited to a High-Nanog state.

Low-Nanog:VNP cells

In Low-Nanog:VNP control cells (DMSO), 15.26% of High-Nanog cells express high levels of *Sox3* transcripts, while 50.20% express high levels of *Car2* transcripts, whereas 52.85% and 81.37% of Low-Nanog cells express high levels of *Sox3* and *Car2* transcripts, respectively.

Correlating the expression of both genes in the same cells show that 13.65% of High-Nanog cells and 48.29% of Low-Nanog cells express high levels of *Fgf5* an *Sox3*. This result indicates that the majority of "primed" cells expressing *Sox3* also express *Car2*.

In mESCs cultured with GSK343 for 48h, there is an increase in the percentage of High-Nanog cells that express high levels of *Sox3* (from 15.26% in DMSO to 19.57% in GSK343) and a decrease in the percentage of High-Nanog cells with high expression of *Car2* (from 50.20% in DMSO to 40.76% in GSK343). Although there is not an alteration in the mean number of transcripts per cell, this single-cell analysis reveals an increased frequency of High-Nanog cells with high levels of *Sox3*. This indicates that the inhibition of PRC2 activity by GSK343, and consequent non-methylation of H3K27 in the genome region controlling *Sox3*, results in higher expression in High-Nanog cells.

Compared to GSK343, AA has an opposite effect in cells that transit to a High-Nanog state, with a decrease in the percentage of new High-Nanog cells expressing high levels of *Sox3* (8.46% versus 15.26% in DMSO). This supports our hypothesis that increased Tet1 activity and associated higher PRC2-mediated H3K27me3 deposition, acts downstream of Nanog to repress *Sox3* expression.

To further test our hypothesis, the two modulators were used together. Our prediction being that the AA effect on increasing repression of priming gene expression in new High-Nanog cells should be dependent on PRC2 activity. The results show that AA+GSK343 treatment leads to a marked increase in the percentage of new High-Nanog cells that also express high levels of *Sox3* mRNA (25.00%, versus 8.46% in AA and 15.26% in DMSO). The repressive effect of AA is therefore completely reverted by inhibition of H3K27 methylation by GSK343. This result fully supports the proposed model of *Sox3* repression by the combined action of Nanog/Tet1/PRC2.

Table 4.10 - Percentage of mESCs with High Expression of Priming Genes Relatively to Nanog Level. Low- and High-Nanog defined by mRNA expression, and the percentage of High-Nanog cells measured by mRNA and Nanog:VNP.

		% High-Car2		%High-Sox3		% High-Double		% High-Nanog (mRNA)	% High-Nanog (VNP)
		Low-Nanog	High-Nanog	Low-Nanog	High-Nanog	Low-Nanog	High-Nanog		
Low-Nanog:VNP	DMSO	81.37	50.20	52.85	15.26	48.29	13.65	48.63	14.6
	GSK343	73.97	40.76	46.58	19.57	41.10	15.76	55.76	15.4
	AA	84.00	48.26	29.00	8.46	28.00	8.46	66.78	49.9
	AA+GSK343	83.94	47.84	23.36	25.00	21.90	14.66	62.87	52.7
High-Nanog:VNP	DMSO	59.82	37.08	36.61	16.48	26.79	10.11	70.45	55.6
	GSK343	51.02	36.09	23.47	7.89	18.37	6.02	73.08	67.4
	AA	60.53	26.46	18.42	2.75	15.79	1.72	88.45	73.8
	AA+GSK343	66.67	37.56	23.33	4.52	20.00	3.62	88.05	81.2

5. Discussion and Conclusions

This project aimed to assess a previously proposed model of priming gene regulation in mESCs by Nanog, Tet1 and PRC2. In this model, Nanog recruits Tet1 to the regulatory regions of priming genes, with Tet1 catalysing the conversion of 5mC into 5hmc, increasing the hypomethylation of these regions. These hypomethylated regions are then recognized by PRC2, which will methylate surrounding nucleosomes on H3K27, creating a repressive environment that thwarts priming gene expression.

This model has its origins on the observations that Nanog exhibits a highly heterogeneous expression in mESCs, with some cells exhibiting high expression (High-Nanog) and others exhibiting very low or null expression (Low-Nanog). This heterogeneity is accompanied by molecular and functional differences, with High-Nanog cells being considered to be in a pristine state of pluripotency, while Low-Nanog cells are proposed to be in a state of primed pluripotency (Martinez Arias and Brickman, 2011). In High-Nanog cells, the pluripotency network is fully operational, being maintained by the core transcription factors Oct4, Sox2 and Nanog. In Low-Nanog cells, Oct4 and Sox2 are able to maintain the pluripotency network active, but genes usually involved in lineage-choice and differentiation begin to be expressed.

Previous work of the laboratory led to the identification of genes exclusively expressed in the Low-Nanog state, termed priming genes. As these genes are mainly expressed when Nanog is not present, they are candidates to be repressed by it. An *in silico* chromatin enrichment analysis was also performed, indicating that these genes are mainly bound by components of PRC2, a complex involved in gene repression.

Through smRNA-FISH, it was possible to quantify the exact number of mRNAs in each cell, correlating *Nanog* expression with that of priming genes. Besides the single-cell information, the analysis of a great number of cells allow the quantification of the population dynamics. Through this method, the expression of *Nanog* and priming genes (*Car2*, *Crabp2*, *Dnmt3b*, *Fgf5* and *Sox3*) was quantified.

First, expression of *Nanog* and *Sox2* (pluripotency genes) in Serum/LIF was assessed by smRNA-FISH. This experiment showed that while *Sox2* can be considered to be expressed homogeneously, as shown by the Gaussian-like distribution of *Sox2* mRNA in the cell population, *Nanog* is expressed very heterogeneously, with most cells showing low expression and some cells with very high expression. As expected, expression of *Nanog* and *Sox2* is positively correlated, although 30% of mESCs expressing high levels of *Sox2* and low levels of *Nanog*, likely corresponding to cells in the primed state of pluripotency.

Next, a detailed smRNA-FISH expression study was performed on various priming genes, namely *Car2*, *Crabp2*, *Dnmt3b*, *Fgf5* and *Sox3*, in mESCs cultured in Serum/LIF conditions. These experiments revealed that these priming genes present long tail distributions of mRNA expression, with low expression in the majority of cells and higher expression in a small fraction of cells, and that high expression occurs mainly in cells with low *Nanog* expression, confirming the existence of priming.

Definition of what is high gene expression, in terms of transcript counts, is necessary for the classification of mESCs into different states. Gene expression is a very dynamic process and many genes are expressed at low levels in mESCs, as it is the case of priming genes. Since priming genes are repressed in High-Nanog cells and transcribed in Low-Nanog cells, one can obtain the two Nanog

populations by FACS, using a mESC line with a fluorescent Nanog reporter (Nd mESCs). Analysis of gene expression by smRNA-FISH on these cells confirmed that Low-Nanog:VNP cells express higher levels of the priming genes *Fgf5* and Sox3 than High-Nanog cells. As these cells still show a very low expression of priming genes, we used this information to define a threshold separating high from low expression levels, assuming that the observed expression values in High-Nanog cells represents the limit of low expression. These thresholds were then applied in our expression analysis of mESCs cultured in Serum/LIF mESCs, with approximately 10% of cells being classified as having high *Fgf5* or *Sox3* expression. By comparison, similar thresholds (approximately 10% of mESCs with high expression) were defined for the other priming genes.

After describing the normal expression of each gene in mESCs grown in Serum/LIF conditions, and defining a threshold to separate high expression of each gene, it was time to test the proposed model for priming gene regulation by Nanog, Tet1 and PRC2.

To test the proposed model, two small molecule modulators were used. Ascorbic acid (AA), which is a known stimulator of Tet1 activity, and GSK343, which inhibits the activity of Ezh2, the catalytical subunit of PRC2. The model predicts that GSK343 treatment of mESCs should lead to an increase in expression of priming genes, by inhibiting gene repression caused by PRC2-mediated H3K27me3. The opposite effect should be produced from AA, decreasing the expression of priming genes by promoting CpG demethylation by Tet1, recruitment of PRC2, and deposition of H3K27me3 around priming gene regulatory regions. As a further test to the model, we exposed mESCs to the two compounds simultaneously, to evaluate whether priming gene repression by AA is mediated by PRC2. If indeed PRC2 is acting downstream of Tet1, we predicted that decreased priming gene expression caused by AA should be reversed.

Our smRNA-FISH results allowed us to directly correlate Nanog expression with priming gene expression in the different conditions, to test our model predictions. When mESCs were exposed to AA for 48h, we observed an increased frequency of cells with high *Car2* and *Crabp2* expression in both High- and Low-Nanog cells. This indicates that a decrease in DNA methylation leads to increased expression of these genes. When GSK343 was added to AA, this increase was reversed, again in both Low- and High-Nanog sates, revealing that *Car2* and *Crabp2* are regulated by the Tet1/PRC2 module, but in a manner that is independent of Nanog.

Fgf5 and *Sox3* behave differently from *Car2* and *Crabp2*, in that expression of these genes is increased by GSK343, suggesting that expression is indeed regulated by PRC2. While *Sox3* regulation is Nanog-dependent, *Fgf5* seem to be regulated in a Nanog-independent manner, increasing its expression in both Nanog states. When mESCs were cultured in the presence of AA, expression of *Fgf5* and *Sox3* was repressed, indicating that they are regulated by Tet1. However, simultaneous exposure to AA and GSK343 causes a reversion of *Fgf5* repression by AA, while *Sox3* expression was similar to AA alone. This *Sox3* result contrasts with previous results obtained in the laboratory, in which *Sox3* AA-mediated repression was indeed reverted by GSK343 addition. At this moment, we cannot explain this discrepancy. One possible explanation was that bulk cultures of mESCs contain both Low- and High-Nanog expressing cells, which can fluctuate rapidly between states and mask possible effects of the modulators. We then used purified populations of Low- and High-Nanog cells, which were treated separately with AA, GSK343 or AA+GSK343, for 48h. In this experiment *Nanog*, *Car2* and *Sox3*

expressions were analysed. The results revealed that AA causes an increase of *Nanog* expression in Low-Nanog:VNP cells, as quantified by mRNA and Nanog:VNP. AA also acted on *Sox3*, causing a decrease of expression, which is partially rescued when cells were treated with both modulators. This result supports the proposed model and clarifies the previous experiment, in which the treatment with both modulators did not increased *Sox3* expression.

When analysing *Car2* and *Sox3* expression in Low-Nanog:VNP mESCs that transited to High-Nanog during the 48h of the experiment, the results reveal that GSK343 reduces the expression of *Car2*, indicating that *Car2* is Ezh2/PRC2 regulated, independently of Nanog. *Sox3*, on the other hand, is regulated by Nanog/Tet1/PRC2, with GSK343 increasing *Sox3* expression in the High-Nanog ESCs, and rescuing AAs' decrease of expression, when ESCs were cultured with both modulators.

There results corroborate our proposed model, in which Nanog regulates priming gene expression, with PRC2 acting downstream of Tet1 to repress priming genes like Sox3 (Figure 4.10).

It is important to notice that in the experiments with Low- and High-Nanog:VNP cells treated with the modulators the number of cells was relatively lower than all other experiments. It is necessary to increase the number of analysed cells to confirm these results.

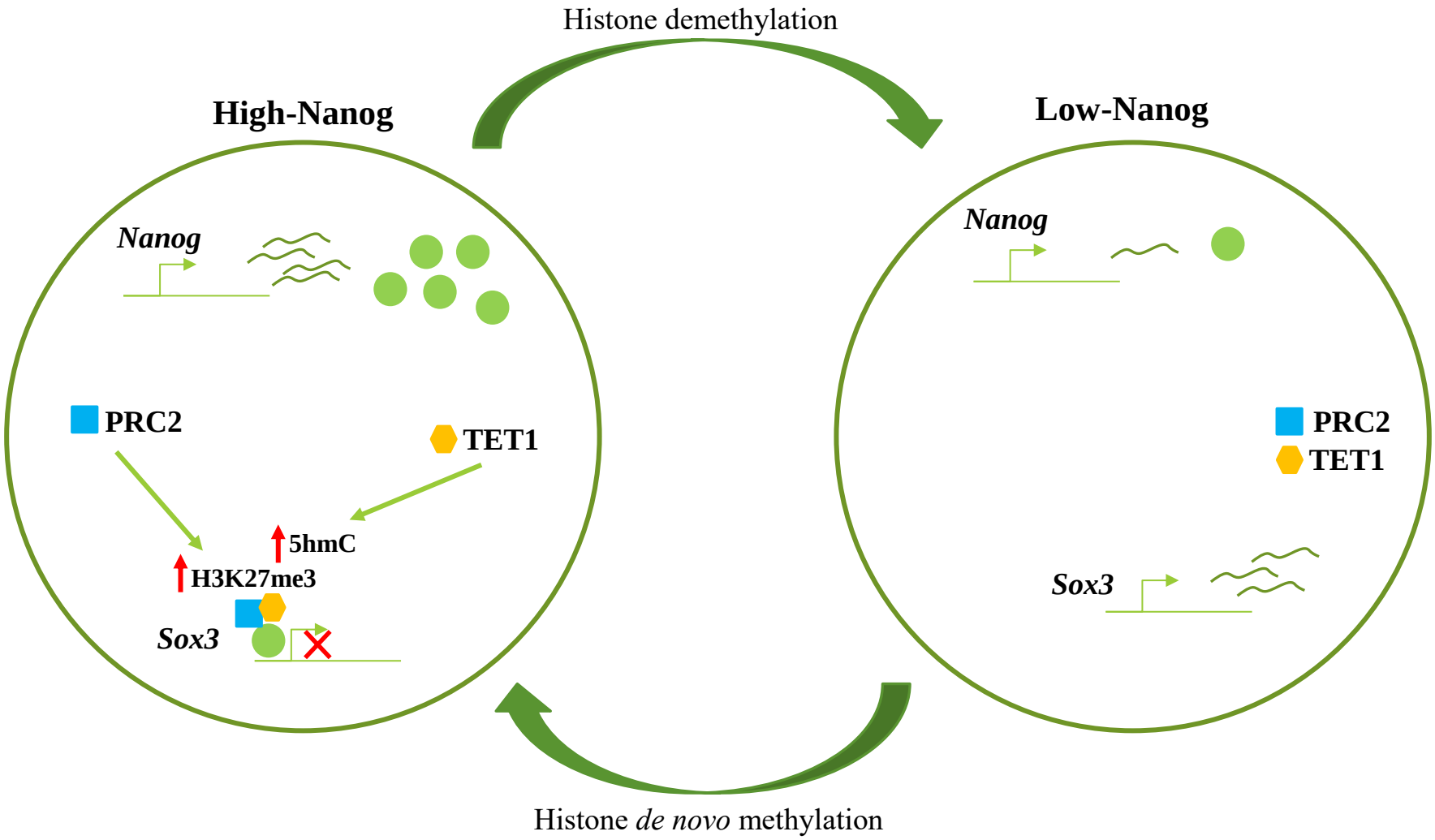

Figure 4.10 – Model of *Sox3* regulation by Nanog, Tet1 and PRC2. In High-Nanog mESCs, Nanog binds to regulatory regions of *Sox3* and recruits Tet1, which catalyses the conversion of 5mC into 5hmC, generating hypomethylated regions. These regions are then recognized by PRC2, that trimethylates H3K27, ensuring repression of *Sox3* expression.

5.1 Future Perspectives

In this project, five priming genes were analysed to test the proposed model of gene regulation by Nanog. It is, therefore, critical to extend the number of genes analysed in order to support further the general model, and/or understand why this model might be only valid for some priming genes.

If more genes validate our hypothesis, the next step shall be the characterization of the chromatin status of priming gene promoters, in the various experimental conditions. This shall aim to correlate the observed alterations in priming gene expression with changes in 5hmC and H3K27me3 levels at the corresponding promoters, caused by interfering with Tet1 and PRC2 activities by the modulators. Chromatin and methyl-DNA immunoprecipitation with H3k27me3 and 5hmC specific antibodies, followed by sequencing (MeDIP and ChIP-Seq; Weber et al., 2005; Schmidt et al., 2009;), will identify all the genomic regions with these modifications.

Here, expression state of Nanog in mESC was defined as having more or less than 50 mRNA molecules. mRNA expression is therefore being used as a proxy for protein expression, which is the molecule with the repressing activity, by binding to DNA and recruiting Tet1. The ideal technique would be one that allows quantification of single mRNAs and single proteins in the same cell, which is not available yet. Very recently, a variation of smRNA-FISH protocol was developed that allows concomitant quantification of mRNA and protein through flow cytometry, called FISH-Flow (Arrigucci et al., 2017). This technique uses the same type of fluorescent probes as smRNA-FISH, besides antibodies to detect protein, and could be used in the future to address the questions here discussed, with more comprehensive results.

The modulators here employed affect the entire genome and are not useful to evaluate the model at the single genomic location. To achieve this, the ideal method would be to direct chromatin modifiers to specific genes/genomic regions. This technology is becoming available, using variations of CRISPR/Cas9 methods. One can now envisage to direct Tet1 and PRC2 activities to the promoter of specific priming genes and therefore test our model in a fine manner. This will require the use of guide RNAs targeted to specific regions of the chosen priming gene regulatory region(s), coupled with modified Cas9 proteins that contain fusions with Ezh2 or Tet1 (Liu et al., 2016). This would allow us to drive the specific addition of 5hmC or H3K27me3 modifications to specific gene promoters and thereby manipulate their activity. In this way, it will be possible to directly modify the chromatin environment of single priming genes and test with great precision the predictions of our model on how Nanog regulates priming genes in pluripotency.

Genes associated with lineage commitment are usually associated with bivalent chromatin, characterized by H3K27me3 and H3K4me3. It would be interesting to test if the inhibition of H3K4me3 deposition would lead to repression of genes. Following a similar approach to the one here employed, it is possible to inhibit MLL1 activity with the chemical modulator MM-401 (Cao et al., 2014).

6. References

Abdelalim, E. M., Emara, M. M. and Kolatkar, P. R. (2014). The SOX Transcription Factors as Key Players in Pluripotent Stem Cells. *Stem Cells Dev.* **23**, 2687–2699.

Abranches, E., Bekman, E. and Henrique, D. (2013). Generation and Characterization of a Novel Mouse Embryonic Stem Cell Line with a Dynamic Reporter of Nanog Expression. *PLoS One* **8**, 1–13.

Abranches, E., Guedes, A. M. V, Moravec, M., Maamar, H., Svoboda, P., Raj, A. and Henrique, D. (2014). Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. *Development* **141**, 2770–9.

Acampora, D., Di Giovannantonio, L. G. and Simeone, A. (2013). Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development* **140**, 43–55.

Aloia, L., Di Stefano, B. and Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. *Development* **140**, 2525–2534.

Ang, Y. S., Tsai, S. Y., Lee, D. F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183–187.

Armstrong, L., Hughes, O., Yung, S., Hyslop, L., Stewart, R., Wappler, I., Peters, H., Walter, T., Stojkovic, P., Evans, J., et al. (2006). The role of PI3K/AKT, MAPK/ERK and NFκβ signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis. *Hum. Mol. Genet.* **15**, 1894–1913.

Arrigucci, R., Bushkin, Y., Radford, F., Lakehal, K., Vir, P., Pine, R., Martin, D., Sugarman, J., Zhao, Y., Yap, G. S., et al. (2017). FISH-Flow, a protocol for the concurrent detection of mRNA and protein in single cells using fluorescence in situ hybridization and flow cytometry. *Nat. Protoc.* **12**, 1245–1260.

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development on SOX2 function. *Genes Dev.* **17**, 126–140.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8**, 532–8.

Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C. and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* **280**, 17732–17736.

Bartolomei, M. S. and Ferguson-Smith, A. C. (2011). Mammalian genomic imprinting. *Cold Spring Harb. Perspect. Biol.* **3**, 1–17.

Beddington, R. S. and Robertson, E. J. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**, 733–737.

Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T. and Schreiber, S. L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci.* **99**, 8695–8700.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315–326.

Blaschke, K., Ebata, K. T., Karimi, M. M., Zepeda-, J. a, Goyal, P., Mahapatra, S., Tam, A., Laird, D. J., Rao, A., Lorincz, M. C., et al. (2013). Vitamin C induces Tet-dependent DNA demethylation in ESCs to promote a blastocyst-like state. *Nature* **500**, 222–226.

Boeuf, H., Hauss, C., Graeve, F. D., Baran, N. and Kedinger, C. (1997). Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. *J. Cell Biol.* **138**, 1207–17.

Boroviak, T., Loos, R., Bertone, P., Smith, A. and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* **16**, 516–528.

Braude, P., Bolton, V. and Moore, S. (1988). Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* **332**, 459–461.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. and Zhiang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb group silencing. *Science (80-.).* **298**, 1039–1044.

Cao, F., Townsend, E. C., Karatas, H., Xu, J., Li, L., Lee, S., Liu, L., Chen, Y., Ouillette, P., Zhu, J., et al. (2014). Targeting MLL1 H3K4 Methyltransferase Activity in Mixed-Lineage Leukemia. *Mol. Cell* **53**, 247–261.

Catena, R., Tiveron, C., Ronchi, A., Porta, S., Ferri, A., Tatangelo, L., Cavallaro, M., Favaro, R., Ottolenghi, S., Reinbold, R., et al. (2004). Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells. *J. Biol. Chem.* **279**, 41846–41857.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L. and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–4.

Chazaud, C., Yamanaka, Y., Pawson, T. and Rossant, J. (2006). Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. *Dev. Cell* **10**, 615–624.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., et al. (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell* **133**, 1106–1117.

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G., Clark, N. R. and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128.

Chen, C. Y., Cheng, Y. Y., Yen, C. Y. T. and Hsieh, P. C. H. (2017). Mechanisms of pluripotency maintenance in mouse embryonic stem cells. *Cell. Mol. Life Sci.* **74**, 1805–1817.

Chew, J., Loh, Y., Zhang, W., Chen, X., Tam, W., Yeap, L., Li, P., Ang, Y., Robson, P., Ng, H., et al. (2006). Reciprocal Transcriptional Regulation of Complex in Embryonic Stem Cells Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4 / Sox2 Complex in Embryonic Stem Cells. *Mol. Cell. Biol.* **25**, 6031–6046.

Choi, J., Huebner, A. J., Clement, K., Walsh, R. M., Savol, A., Lin, K., Gu, H., Di Stefano, B., Brumbaugh, J., Kim, S.-Y., et al. (2017). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **1**, 1–7.

Costa, Y., Ding, J., Theunissen, T. W., Faiola, F., Hore, T. a, Shliaha, P. V, Fidalgo, M., Saunders,

A., Lawrence, M., Dietmann, S., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495**, 370–4.

Crosby, I. M., Gandolfi, F. and Moor, R. M. (1988). Control of protein synthesis during early cleavage of sheep embryos. *J. Reprod. Fertil.* **82**, 769–775.

Cross, M. H. (1973). Active Sodium and Chloride Transport Across The Rabbit Blastocoele Wall. *Biol. Reprod.* **8**, 566–575.

Dar, R. D., Razooky, B. S., Singh, A., Trimeloni, T. V, McCollum, J. M., Cox, C. D., Simpson, M. L. and Weinberger, L. S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17454–9.

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H. G. and Stewart, a F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141**, 526–37.

Dou, Y., Milne, T. a, Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D. and Roeder, R. G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat. Struct. Mol. Biol.* **13**, 713–719.

Ducibella, T. and Anderson, E. (1975). Cell shape and membrane changes in the eight-cell mouse embryo: Prerequisites for morphogenesis of the blastocyst. *Dev. Biol.* **47**, 45–58.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D. J., Dash, C., Bazett-Jones, D. P., Le Grice, S., McKay, R. D. G., Buetow, K. H., et al. (2008). Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell* **2**, 437–447.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156.

Faust, C., Lawson, K. A., Schork, N. J., Thiel, B. and Magnuson, T. (1998). The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development* **125**, 4495–4506.

Fleming, T. P. (1987). A quantitative analysis of cell allocation to trophectoderm and inner cell mass in the mouse blastocyst. *Dev. Biol.* **119**, 520–531.

Furusawa, T., Ohkoshi, K., Honda, C., Takahashi, S. and Tokunaga, T. (2004). Embryonic stem cells expressing both platelet endothelial cell adhesion molecule-1 and stage-specific embryonic antigen-1 differentiate predominantly into epiblast cells in a chimeric embryo. *Biol. Reprod.* **70**, 1452–1457.

Gagliardi, A., Mullin, N. P., Ying Tan, Z., Colby, D., Kousa, A. I., Halbritter, F., Weiss, J. T., Felker, A., Bezstarosti, K., Favaro, R., et al. (2013). A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *Embo J* **32**, 2231–2247.

Graf, T. and Stadtfeld, M. (2008). Heterogeneity of Embryonic and Adult Stem Cells. *Cell Stem Cell* **3**, 480–483.

Graham, S. J. L. and Zernicka-Goetz, M. (2016). *The Acquisition of Cell Fate in Mouse Development: How Do Cells First Become Heterogeneous?* 1st ed. Elsevier Inc.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610.

Habibi, E., Brinkman, A. B., Arand, J., Kroeze, L. I., Kerstens, H. H. D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N. C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell*

13, 360–369.

Handyside, a H. (1978). Time of commitment of inside cells isolated from preimplantation mouse embryos. *J. Embryol. Exp. Morphol.* **45**, 37–53.

Handyside, A. H. and Hunter, S. (1986). Cell division and death in the mouse blastocyst before implantation. *Roux's Arch. Dev. Biol.* **195**, 519–526.

Hayashi, K., Lopes, S. M. C. de S., Tang, F. and Surani, M. A. (2008). Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. *Cell Stem Cell* **3**, 391–401.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science (80-.).* **333**, 1303–1307.

Hébert, J. M., Boyle, M. and Martin, G. R. (1991). mRNA localization studies suggest that murine FGF-5 plays a role in gastrulation. *Development* **112**, 407–415.

Hermann, A., Goyal, R. and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J. Biol. Chem.* **279**, 48350–48359.

Hillman, N., Sherman, M. I. and Graham, C. (1972). The effect of spatial arrangement on cell determination during mouse development. *Embryol. exp. Morph* **28**, 263–278.

Hirate, Y., Hirahara, S., Inoue, K. I., Suzuki, A., Alarcon, V. B., Akimoto, K., Hirai, T., Hara, T., Adachi, M., Chida, K., et al. (2013). Polarity-dependent distribution of angiomotin localizes hippo signaling in preimplantation embryos. *Curr. Biol.* **23**, 1181–1194.

Hore, T. A., von Meyenn, F., Ravichandran, M., Bachman, M., Ficz, G., Oxley, D., Santos, F., Balasubramanian, S., Jurkowski, T. P. and Reik, W. (2016). Retinol and ascorbate drive erasure of epigenetic memory and enhance reprogramming to naïve pluripotency by complementary mechanisms. *Proc. Natl. Acad. Sci.* **113**, 12202–12207.

Huang, Y., Osorno, R., Tsakiridis, A. and Wilson, V. (2012). In Vivo Differentiation Potential of Epiblast Stem Cells Revealed by Chimeric Embryo Formation. *Cell Rep.* **2**, 1571–1578.

Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., Zepeda-Martinez, J. A., Pape, U. J., Jacobsen, S. E., Peters, B., et al. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Natl. Acad. Sci.* **111**, 1361–1366.

Hyafil, F., Babinet, C. and Jacob, F. (1981). Cell-cell interactions in early embryogenesis: a molecular approach to the role of calcium. *Cell* **26**, 447–454.

Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C. and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133.

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C. and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science (80-.).* **333**, 1300–1303.

Johnson, M. H. and Ziomek, C. A. (1981). The foundation of two distinct cell lineages within the mouse morula. *Cell* **24**, 71–80.

Kanter, I. and Kalisky, T. (2015). Single cell transcriptomics: methods and applications. *Front. Oncol.* **5**, 53.

Keller, G. M. (1995). In vitro differentiation of embryonic stem cells. *Curr. Opin. Cell Biol.* **7**, 862–869.

Kelly, K. F., Ng, D. Y., Jayakumaran, G., Wood, G. A., Koide, H. and Doble, B. W. (2011). β-catenin enhances Oct-4 activity and reinforces pluripotency through a TCF-independent mechanism. *Cell Stem Cell* **8**, 214–227.

Kim, J., Chu, J., Shen, X., Wang, J. and Orkin, S. H. (2008). An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. *Cell* **132**, 1049–1061.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell* **128**, 693–705.

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120.

Lanner, F. and Rossant, J. (2010). The role of FGF/Erk signaling in pluripotent cells. *Development* **137**, 3351–3360.

Lee, H. J., Hore, T. A. and Reik, W. (2014). Reprogramming the methylome: Erasing memory and creating diversity. *Cell Stem Cell* **14**, 710–719.

Leung, C. Y. and Zernicka-Goetz, M. (2013). Angiomotin prevents pluripotent lineage differentiation in mouse embryos via Hippo pathway-dependent and -independent mechanisms. *Nat. Commun.* **4**, 1–11.

Liang, G., Lin, J. C. Y., Wei, V., Yoo, C., Cheng, J. C., Nguyen, C. T., Weisenberger, D. J., Egger, G., Takai, D., Gonzales, F. A., et al. (2004). Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc. Natl. Acad. Sci.* **101**, 7357–7362.

Liu, X. S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R. A. and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233–247.e17.

Luger, K., Mäder, a W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260.

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., Firth, A., Singer, O., Trono, D. and Pfaff, S. L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63.

Margueron, R. and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–9.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., Smith, A., et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604.

Marson, A., Levine, S. S., Cole, M. F., Frampton, G. M., Brambrink, T., Johnstone, S., Guenther, M. G., Johnston, W. K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells. *Cell* **134**, 521–533.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634–7638.

Martin Gonzalez, J., Morgani, S. M., Bone, R. A., Bonderup, K., Abelchian, S., Brakebusch, C. and Brickman, J. M. (2016). Embryonic Stem Cell Culture Conditions Support Distinct States Associated with Different Developmental Stages and Potency. *Stem Cell Reports* **7**, 177–191.

Martinez Arias, A. and Brickman, J. M. (2011). Gene expression heterogeneities in embryonic stem

cell populations: Origin and function. *Curr. Opin. Cell Biol.* **23**, 650–656.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**, 625-U26.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P. J., Brown, D. T. and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105–116.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S. (2003). The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642.

Morris, S. A., Teo, R. T. Y., Li, H., Robson, P., Glover, D. M. and Zernicka-Goetz, M. (2010). Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proc. Natl. Acad. Sci.* **107**, 6364–6369.

Mullin, N. P., Yates, A., Rowe, A. J., Nijmeijer, B., Colby, D., Barlow, P. N., Walkinshaw, M. D. and Chambers, I. (2008). The pluripotency rheostat Nanog functions as a dimer. *Biochem. J.* **411**, 227–31.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H. and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo dependes on the POU transcription factor Oct4. *Cell* **95**, 379–391.

Nicolas, D., Phillips, N. E. and Naef, F. (2017). What shapes eukaryotic transcriptional bursting? *Mol. BioSyst.* **13**, 1280–1290.

Nightingale, K. P. (2015). Epigenetics - What it is ans Why it matters. *Physician Exec.* **24**, 43–45.

Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K. and Sasaki, H. (2008). Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech. Dev.* **125**, 270–283.

Nishioka, N., Inoue, K. ichi, Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R. O., Ogonuki, N., et al. (2009). The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass. *Dev. Cell* **16**, 398–410.

Niwa, H., Burdon, T., Chambers, I., Niwa, H., Burdon, T., Chambers, I. and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. 2048–2060.

Niwa, H., Miyazaki, J. and Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372–376.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R. and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917–929.

Niwa, H., Ogawa, K., Shimosato, D. and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460**, 118–122.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S. C., Surani, M. A. and Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. *Mol. Cell. Biol.* **21**, 4330–4336.

Ochiai, H., Sugawara, T., Sakuma, T. and Yamamoto, T. (2015). Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. *Sci. Rep.* **4**, 7125.

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b

are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257.

Orkin, S. H. and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. *Cell* **145**, 835–850.

Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P. J., Chambers, I. and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development* **139**, 2288–2298.

Palmieri, S. L., Peter, W., Hess, H. and Schöler, H. R. (1994). Oct-4 Transcription Factor Is Differentially Expressed in the Mouse Embryo during Establishment of the First Two Extraembryonic Cell Lineages Involved in Implantation. *Dev. Biol.* **166**, 259–267.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G. A., Stewart, R. and Thomson, J. A. (2007). Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. *Cell Stem Cell* **1**, 299–312.

Papatsenko, D., Darr, H., Kulakovskiy, I. V., Waghray, A., Makeev, V. J., Macarthur, B. D. and Lemischka, I. R. (2015). Single-Cell Analyses of ESCs Reveal Alternative Pluripotent Cell States and Molecular Mechanisms that Control Self-Renewal. *Stem Cell Reports* **5**, 207–220.

Pasini, D., Bracken, A. P., Jensen, M. R., Denchi, E. L. and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* **23**, 4061–4071.

Perona, R. M. and Wassarman, P. M. (1986). Mouse blastocysts hatch in vitro by using a trypsin-like proteinase associated with cells of mural trophectoderm. *Dev. Biol.* **114**, 42–52.

Peyrieiras, N., Hyafil, F., Louvard, D., Ploeght, H. L. and Jacob, F. (1983). Uvomorulin: A nonintegral membrane protein of early mouse embryo. (teratocarcinoma/membranee protein/cell adhesion/Ca2+). *Dev. Biol.* **80**, 6274–6277.

Piko, L. and Clegg, K. B. (1982). Quantitative Changes in Total RNA, Total Poly (A), and Ribosomes in Early Mouse Embryos. *Dev. Biol.* **89**, 362–378.

Raj, A. and Tyagi, S. (2010). *Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes.* 1st ed. Elsevier Inc.

Raj, A. and van Oudenaarden, A. (2009). Single Molecule Approaches to Stochastic Gene Expression. *Annu. Rev. Biophys.* **38**, 255–270.

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–9.

Roeder, I. and Radtke, F. (2009). Stem cell biology meets systems biology. *Development* **136**, 3525–3530.

Rossant, J. and Lis, W. T. (1979). Potential of isolated mouse inner cell masses to form trophectoderm derivatives in vivo. *Dev. Biol.* **70**, 255–261.

Ruberte, E., Friederich, V., Morriss-Kay, G. and Chambon, P. (1992). Differential distribution patterns of CRABP-I and CRABP-II transcripts during mouse embryogenesis. *Development* **115**, 973–989.

Sancho, M., Di-Gregorio, A., George, N., Pozzi, S., Sánchez, J. M., Pernaute, B. and Rodríguez, T. A. (2013). Competitive interactions eliminate unfit embryonic stem cells at the onset of differentiation. *Dev. Cell* **26**, 19–30.

Schmidt, D., Wilson, M. D., Spyrou, C., Brown, G. D., Hadfield, J. and Odom, D. T. (2009). ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240–248.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* **473**, 337–342.

Shi, Y. and Massagué, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685–700.

Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I. and Smith, A. (2009). Nanog Is the Gateway to the Pluripotent Ground State. *Cell* **138**, 722–737.

Singh, U., Sun, T., Shi, W., Schulz, R., Nuber, U. A., Varanou, A., Hemberger, M. C., Elliott, R. W., Ohta, H., Wakayama, T., et al. (2005). Expression and functional analysis of genes deregulated in mouse placental overgrowth models: Car2 and Ncam1. *Dev. Dyn.* **234**, 1034–1045.

Singh, A. M., Hamazaki, T., Hankowski, K. E. and Terada, N. (2007). A Heterogeneous Expression Pattern for Nanog in Embryonic Stem Cells. *Stem Cells* **25**, 2534–2542.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688–690.

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A. and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344.

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *Am. J. Psychol.* **15**, 72.

Spindle, A. I. (1978). Trophoblast regeneration by inner cell masses isolated from cultured mouse embryos. *J. Exp. Zool.* **203**, 483–489.

Strumpf, D. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093–2102.

Surani, M. A., Hayashi, K. and Hajkova, P. (2007). Genetic and Epigenetic Regulators of Pluripotency. *Cell* **128**, 747–762.

Takahashi, K. and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676.

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K. and Surani, M. A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-seq analysis. *Cell Stem Cell* **6**, 468–478.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909–918.

van den Berg, D. L. C., Zhang, W., Yates, A., Engelen, E., Takacs, K., Bezstarosti, K., Demmers, J., Chambers, I. and Poot, R. A. (2008). Estrogen-Related Receptor Beta Interacts with Oct4 To Positively Regulate Nanog Gene Expression. *Mol. Cell. Biol.* **28**, 5986–5995.

Verma, S. K., Tian, X., Lafrance, L. V., Duquenne, C., Suarez, D. P., Newlander, K. A., Romeril, S. P., Burgess, J. L., Grant, S. W., Brackley, J. A., et al. (2012). Identification of potent, selective, cell-Active inhibitors of the histone lysine methyltransferase EZH2. *ACS Med. Chem. Lett.* **3**, 1091–1096.

Villasante, A., Piazzolla, D., Li, H., Gomez-Lopez, G., Djabali, M. and Serrano, M. (2011). Epigenetic regulation of Nanog expression by Ezh2 in pluripotent stem cells. *Cell Cycle* **10**, 1488–1498.

Wagers, A. J. and Weissman, I. L. (2004). Plasticity of adult stem cells. *Cell* **116**, 639–648.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T. W. and Orkin, S. H. (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364–368.

Wang, J., Levasseur, D. N. and Orkin, S. H. (2008). For the article "'Requirement of Nanog dimerization for stem cell self-renewal and pluripotency,'" by Jianlong Wang, Dana N. Levasseur, and Stuart H. Orkin, which appeared in issue 17, April 29, 2008, of. *Pnas* **105**, 6326–6331.

Watson, A. J. and Kidder, G. M. (1988). Immunofluorescence assessment of the timing of appearance and cellular distribution of Na/K-ATPase during mouse embryogenesis. *Dev. Biol.* **126**, 80–90.

Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L. and Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**, 853–862.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. and Feinberg, A. P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat. Genet.* **41**, 246–250.

Wickham, H. (2011). Ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185.

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. a, Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. a and Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684–687.

Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K. and Zhao, K. (2011). Genome-wide analysis of distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* **25**, 679–684.

Yagi, M., Kishigami, S., Tanaka, A., Semi, K., Mizutani, E., Wakayama, S., Wakayama, T., Yamamoto, T. and Yamada, Y. (2017). Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature* 20–22.

Yamaguchi, T. P. and Rossant, J. (1995). Fibroblast growth factors in mammalian development. *Curr Opin Genet Dev* **5**, 485–491.

Yamanaka, Y., Lanner, F. and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715–724.

Yin, R., Mao, S. Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., et al. (2013). Ascorbic acid enhances tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J. Am. Chem. Soc.* **135**, 10396–10403.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523.

Zernicka-Goetz, M. (2005). Cleavage pattern and emerging asymmetry of the mouse embryo. *Nat. Rev. Mol. Cell Biol.* **6**, 919–928.

Zhang, P., Andrianakos, R., Yang, Y., Liu, C. and Lu, W. (2010). Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression. *J. Biol. Chem.* **285**, 9180–9189.

Zhao, X. D., Han, X., Chew, J. L., Liu, J., Chiu, K. P., Choo, A., Orlov, Y. L., Sung, W. K., Shahab, A., Kuznetsov, V. A., et al. (2007). Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells. *Cell Stem Cell* **1**, 286–298.

Ziomek, C. A. and Johnson, M. H. (1980). Cell surface interaction induces polarization of mouse 8-cell blastomeres at compaction. *Cell* **21**, 935–942.

7. Supplementary Materials

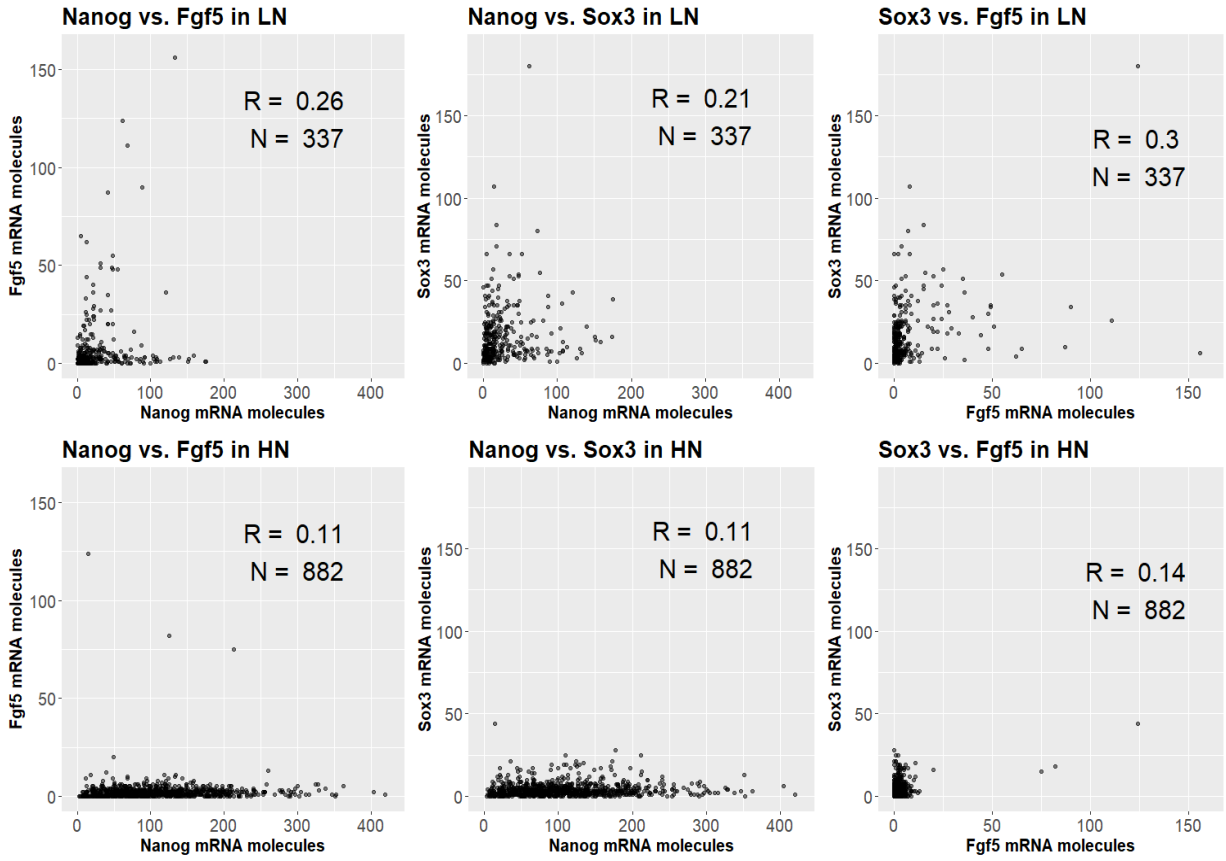

Figure 7.1 – Dispersion Plots of Low- and High-Nanog:VNP mESCs. Dispersion plots with the correlation between each of the genes analysed: *Fgf5*, *Nanog* and *Sox3* in Low-Nanog:VNP (LN) and High-Nanog:VNP (HN). R – Spearman correlation coefficient, N – number of cells analysed. Each dot represents one mESC.

Table 7.1 – Statistical Measurements for *Fgf5*, *Nanog* and *Sox3* in Low- and High-Nanog:VNP mESCs. For all genes, mean, median, standard deviation (SD), variance, Fano factor (FF), coefficient of variation (CV), minimum (Min) and maximum (Max) mRNA counts and number of mESCs analysed (N) are presented.

	Gene	Mean	Median	SD	Variance	FF	CV	Min	Max	N
Low-Nanog:VNP	Fgf5	7.26	2.0	16.93	286.54	39.45	2.33	0	156	337
	Nanog	26.33	13.0	32.39	1048.87	39.83	1.23	0	175	
	Sox3	16.15	11.0	17.76	315.50	19.53	1.10	0	180	
High-Nanog:VNP	Fgf5	2.34	2.0	5.85	34.18	14.61	2.50	0	124	882
	Nanog	107.79	96.5	69.20	4788.97	44.43	0.64	3	419	
	Sox3	3.92	3.0	4.11	16.90	4.31	1.05	0	44	

Table 7.2 – Defined Thresholds for Priming Genes. The threshold used and the correspondent percentage of High-expressing cells, in Serum/LIF.

	Fgf5	Sox3	Car2	Crabp2	Dnmt3b
Threshold	10	20	100	20	15
% High-expressing cells	10.51	12.70	9.66	10.04	11.43

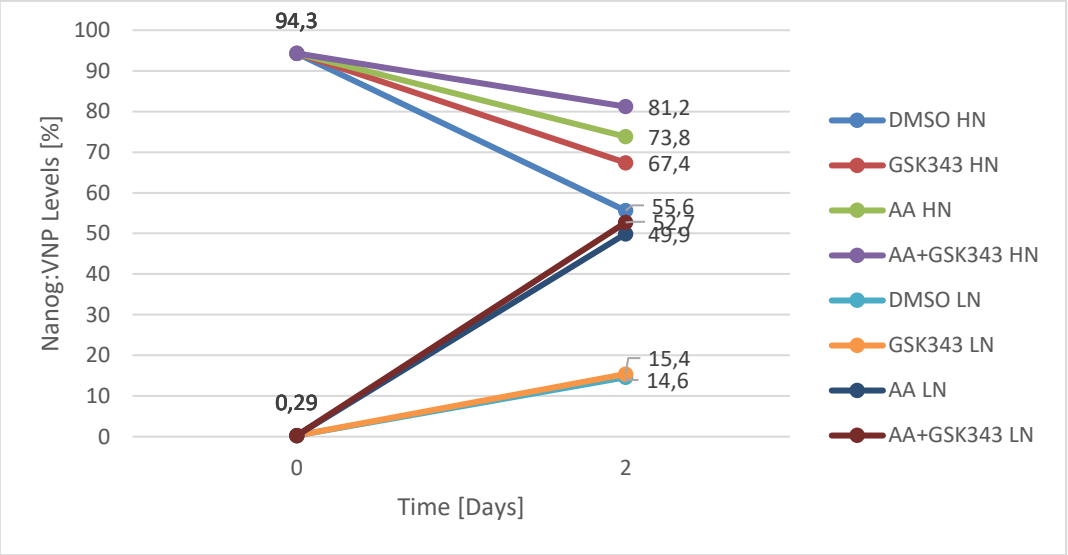

Figure 7.2 – Nanog:VNP expression in Low- and High-Nanog:VNP mESCs treated with the Chemical Modulators. mESC were cultured for 48h in the presence of modulators and Nanog:VNP was assessed by flow cytometry in the beginning (0 days) and the end (2 days) of the experiment. HN: High-Nanog:VNP; LN: Low-Nanog:VNP; experiment performed in colaboration with Raquel Calçada, a previous student in the laboratory.